Modelling the dependence structure of multivariate and spatial extremes

Boris Beranger

Université Pierre et Marie Curie – Paris 6
University of New South Wales

École Doctorale de Sciences Mathématiques de Paris Centre

Thèse de doctorat

Discipline : Mathématiques Appliquées
Specialité : Statistique

présentée par

Boris BÉRANGER

Modélisation de la structure de dépendance d'extrêmes multivariés et spatiaux

---

dirigée par Michel Broniatowski et Scott Sisson

Soutenue le 18 janvier 2016 devant le jury composé de :

M. Gérard Biau	Université Paris 6	Examinateur
M. Michel Broniatowski	Université Paris 6	Directeur
M. Anthony Davison	École Polytechnique Fédérale de Lausanne	Examinateur
M. Clément Dombry	Université de Franche–Comté	Rapporteur
M. Olivier Lopez	Université Paris 6	Examinateur
M. Simone Padoan	Bocconi University of Milan	Examinateur

School of Mathematics and Statistics
The Red Centre, Centre Wing
Kensington Campus
UNSW Sydney, NSW 2051
Australia

Graduate Research School
Lvl 2 South Wing Rupert Myers Building
Gate 14 Barker Street Entrance
Kensington Campus
UNSW Sydney, NSW 2051
Australia

Laboratoire de Statistique Théorique et Appliquée (LSTA)
UPMC
4, place Jussieu
75252 Paris Cedex 05
Boîte courrier 158

UPMC
Ecole Doctorale de Sciences
Mathématiques de Paris Centre
4 place Jussieu
75252 Paris Cedex 05
Boite courrier 290

Acknowledgements

First and foremost I would like to thank my supervisors Prof. Michel Broniatowski and A/Prof. Scott Sisson as well as my co-supervisor Dr. Simone Padoan.

I am thankful to Prof. Michel Broniatowski for giving me the opportunity to do this joint PhD between University Pierre and Marie Curie, Paris 6 and UNSW Australia. Furthermore his extended knowledge of Mathematics and the precision of his reasoning have pushed me forward and gave me the motivation to achieve this work. I am lucky to have had A/Prof. Scott Sisson as a supervisor. His unconditional support, positive attitude and patience have been greatly appreciated. I would also like to thank Scott for his availability, the weekly Monday morning meetings and most importantly for introducing me to Dr Simone Padoan.

I would also like to thank Clément Dombry and Philippe Naveau for their interest in my work and reviewing this thesis as well as Gérard Biau, Anthony Davison and Olivier Lopez who honoured me by accepting to be members of the jury of my PhD defense.

It has been an extreme(!) pleasure to work with Dr. Simone Padoan. Thank you for sharing your passion for the good work, your investment and our discussions about statistics or otherwise! I am also grateful to Simone, Giulia and Isadora for their warm welcome during my visit to Bocconi University, Milan.

I am deeply thankful to have met Dr. Tarn Duong while at UPMC, who I consider as a friend. I particularly appreciated our collaboration as well as our various discussions and his support; thank you Tarn!

Many thanks to Martin Thompson for his help with the Katana cluster and Sarah Perkins for providing the temperature data.

I have had pleasant times with my fellow PhD students and colleagues at UPMC and UNSW. A special thanks to Philippe for being positive, motivating and a true friend. It has been my pleasure to share the office at UPMC with Mokhtar and Tarn but also to share this experience with Alexis, Amadou, Assia, Baptiste, Benjamin, Cécile, Erwan, Matthieu, Patricia, Soumeya and Svetlana. I am also thankful to Louise and Corinne for their help and for being cheerful.

Gery and Susannah, I am grateful for all the good moments shared at uni and out. Ashish, Austen, Benoît, Francis, Gordana, Guilherme, Isaac, Nicole, Thais, Thomas, Reshma, Stefan, Varun, Yoshi, thank you for making UNSW such an enjoyable work environment!

Finally, I would like to thank my friends, the Darley crew and my family for always being there for me, in particular Alex who suggested I study Mathematics around 9 years ago. Above all, I am deeply grateful to my parents for their never-ending confidence and support, this thesis is for you.

Abstract

Projection of future extreme events is a major issue in a large number of areas including the environment and risk management. Although univariate extreme value theory is well understood, there is an increase in complexity when trying to understand the joint extreme behaviour between two or more variables. Particular interest is given to events that are spatial by nature and which define the context of infinite dimensions. Under the assumption that events correspond marginally to univariate extremes, the main focus is then on the dependence structure that links them.

First, we provide a review of parametric dependence models in the multivariate framework and illustrate different estimation strategies. The spatial extension of multivariate extremes is introduced through max-stable processes. We derive the finite-dimensional distribution of the widely used Brown-Resnick model which permits inference via full and composite likelihood methods.

We then use Skew-symmetric distributions to develop a spectral representation of a wider max-stable model: the extremal Skew-t model from which most models available in the literature can be recovered. This model has the nice advantages of exhibiting skewness and non-stationarity, two properties often held by environmental spatial events. The latter enables a larger spectrum of dependence structures. Indicators of extremal dependence can be calculated using its finite-dimensional distribution.

Finally, we introduce a kernel based non-parametric estimation procedure for univariate and multivariate tail density and apply it for model selection. Our method is illustrated by the example of selection of physical climate models.

Keywords

Extreme value theory, multivariate extremes, max-stable processes, finite-dimensional distributions, angular density, dependence, approximate likelihood, composite likelihood, skewed distributions, exploratory data analysis, kernel estimators.

Résumé

Modélisation de la structure de dépendance d'extrêmes multivariés et spatiaux

La prédiction de futurs évènements extrêmes est d'un grand intérêt dans de nombreux domaines tels que l'environnement ou la gestion des risques. Alors que la théorie des valeurs extrêmes univariées est bien connue, la complexité s'accroît lorsque l'on s'intéresse au comportement joint d'extrêmes de plusieurs variables. Un intérêt particulier est porté aux évènements de nature spatiale, définissant le cadre d'un nombre infini de dimensions. Sous l'hypothèse que ces évènements soient marginalement extrêmes, nous focalisons sur la structure de dépendance qui les lie.

Dans un premier temps, nous faisons une revue des modèles paramétriques de dépendance dans le cadre multivarié et présentons différentes méthodes d'estimation. Les processus max-stables permettent l'extension au contexte spatial. Nous dérivons la loi en dimension finie du célèbre modèle de Brown- Resnick, permettant de faire de l'inférence par des méthodes de vraisemblance ou de vraisemblance composée.

Nous utilisons ensuite des lois asymétriques afin de définir la représentation spectrale d'un modèle plus large : le modèle Extremal Skew-t, généralisant la plupart des modèles présents dans la littérature. Ce modèle a l'agréable propriété d'être asymétrique et non-stationnaire, deux notions présentées par les évènements environnementaux spatiaux. Ce dernier permet un large spectre de structures de dépendance. Les indicateurs de dépendance sont obtenus en utilisant la loi en dimension finie.

Enfin, nous présentons une méthode d'estimation non-paramétrique par noyau pour les queues de distributions et l'appliquons à la sélection de modèles. Nous illustrons notre méthode à partir de l'exemple de modèles climatiques.

Mots-clefs

Théorie des valeurs extrêmes, extrêmes multivariés, processus max-stables, lois de dimension finie, densité angulaire, dépendance, vraisemblance approximée, vraisemblance composée, distributions asymétriques, analyse exploratoire de données, estimateurs à noyau.

Contents

Introduction

The analysis of extremes concentrates on modelling events that have a very small probability of occurrence. It has had a growing interest over the last decades with a wide range of applications which includes risk management, finance, insurance, economics, telecommunications, geology, hydrology, meteorology, environmental research and many others. The theory of univariate extremes consists of looking at the limiting distribution of the normalized maximum of a sequence of independent random variables with common distribution function. The limiting distribution, characterized by Fisher and Tippett (1928) and Gnedenko (1943), is known as the generalized extreme value distribution (GEV). The statistical analysis based on the GEV distribution is referred as the block maxima since it models the maxima of observations that fall within some blocks of time (annual or monthly maxima for example). This is closely linked to a second approach: the peak over threshold approach (POT) which considers all the data that exceeds some fixed high threshold. If the distribution of the rescaled maxima is GEV distributed then the distribution of the exceedances converges to a generalized Pareto distribution (GPD), see Balkema and de Haan (1974) and Pickands (1975).

The probability aspects and the statistical inference of univariate extreme distributions is very well known (see for example Resnick (1987); Embrechts et al. (1997); Coles (2001); Beirlant et al. (2006); de Haan and Ferreira (2006)) however in higher dimensions things are more complicated. Finite and infinite sequences of 'extreme' random variables are discussed by the theory of multivariate extremes and the extremes of stochastic processes such as spatial or temporal extremes. In the multivariate case the definition of an extreme value is not obvious, therefore many valid definitions are possible. From the probability point of view, the componentwise maxima and pointwise maxima are the simple ones which are feasible to treat. Both block maxima and POT approaches have been largely covered in the literature. Different approaches to model exceedances over threshold in a multivariate context have been investigated by Ledford and Tawn (1996); Rootzén and Tajvidi (2006); Falk and Guillou (2008); Ferreira and de Haan (2014) and lead to multivariate GPDs. However, the block maxima approach is our main interest here and we will assume to be in this context for the remainder of this thesis.

The block maxima consists of computing for each variable the maxima per block (of time for example). In the multivariate context, the focus is on the univariate marginal distributions and the dependence structure. Dependence is a fundamental issue as it helps to understand how the extremes in one component relate to the extremes in other components. Since stochastic processes are characterized by their multivariate distribution then for spatial and temporal

sequences of extremes, the focus is also on studying the marginal distributions and dependence structure. Univariate extreme value theory is used to standardize these marginal distributions allowing then to define some measures of the extremal dependence. In the context of multivariate extremes, the dependence structure is described by the spectral measure (see Falk et al. (2011)), the Pickands dependence function (Pickands, 1981), the angular density (de Haan and Resnick, 1977; Resnick, 1987), the exponent function (Buishand, 1984; Coles and Tawn, 1991; Smith, 1990), the stable tail dependence function (Huang, 1992; Drees and Huang, 1998) and the coefficient of tail dependence (Ledford and Tawn, 1996, 1997). Additionally the dependence of spatial extremes can be summarized through a spectral representation (de Haan, 1984; Giné et al., 1990). More recently Schlather (2002) introduced a second spectral characterization and Wang and Stoev (2010) proposed an approach based on extremal stochastic integrals.

Differently to the univariate case, in the multivariate case there is not an unique family of limiting distributions when working with the vector of componentwise maxima. Therefore the dependence structure cannot be estimated straight away from the distribution and requires non-parametric or parametric techniques to do so.

Non-parametric estimators of the dependence summaries are given by Einmahl et al. (2001); Einmahl and Segers (2009); de Carvalho et al. (2013) for the spectral measure, by Drees and Huang (1998); Einmahl et al. (2012); Capéraà and Fougères (2000) for the stable tail dependence function, by Ledford and Tawn (1996); Hill (1975) for the coefficient of tail dependence and recently Marcon et al. (2014) for the Pickands dependence function.

Maximum Likelihood estimation is possible for multivariate extreme value distributions. The derivation of the joint density involves differentiation of the exponent function whose complexity increases with the exploding number of terms. Wadsworth and Tawn (2014) proposed a full likelihood estimation procedure based on the conditional distribution of max-stable processes but other likelihood based estimation methods have also recently appeared. The growing interest in high dimensional or spatial scenarios creates new challenges: the high dimensional distributions of some models are not necessarily known, meaning that the usual maximum likelihood estimation method is not possible, or the maximization procedure is computationally heavy. The very advantageous and appealing composite likelihood methods (Lindsay (1988); see Varin et al. (2011) for a review) have then arisen for spatial extremes, Padoan et al. (2010) being the first to use pairwise composite likelihood in the context of spatial extremes using Gaussian max-stable processes. It was also considered for high dimensional data by Smith and Stephenson (2009); Blanchet and Davison (2011); Davison et al. (2012); Davison and Gholamrezaee (2012). The use of triplewise or higher composite likelihood has been contemplated by Genton et al. (2011) and Huser and Davison (2013) who respectively derived the finite-dimensional distributions of Gaussian and Brown-Resnick max-stable processes and compared the efficiency gains between pairwise and triplewise methods. From these experiments, it was concluded that the benefit from using higher degree composite likelihood is only significant when the process is smooth, which isn't the case for most real spatial phenomena. A tapered version of the pairwise composite likelihood was given by Sang and Genton (2014) who showed that not considering the pairwise component between two location far away from each other can bring a gain of efficiency.

Composite likelihoods have been widely used for spatial extremes but also for multivariate extremes, especially under the copula framework. In the latter it has been shown that multivariate extremes modelling can be done through extreme value and non-extreme value copulas, however closed form of the likelihood don't always exist for extreme value copulas (see Ribatet and Sedki (2013); Sang and Gelfand (2010)). Another class of likelihood based estimation procedures relies on the fact that observations, transformed to pseudo-polar coordinates, above a high threshold are approximately from a Poisson point process and the maximum likelihood estimation is then approximated by maximising the product of angular densities. This method, simply called approximate likelihood, has been considered by Coles and Tawn (1994); Cooley et al. (2010); Engelke et al. (2012).

Bayesian methods are another possibility which requires knowledge about the density, hence Ribatet et al. (2012) adapts the composite likelihood in the Bayesian setting. Similarly Sabourin et al. (2013); Sabourin and Naveau (2014) presents the approximate likelihood in the Bayesian setup.

A contribution of this thesis is to provide some advanced results on the modelling of the dependence of multivariate extremes. Chapter 1 reviews the theory of multivariate extremes and lists the parametric models the most widely encountered in the literature. For each of these models, their construction, main features and an interpretation is provided. The expressions of the exponent function and angular density on the d-dimensional simplex are provided. For the Extremal-t model the angular density was not known and is derived for sub-spaces of the simplex. This allows us to derive measures of the extremal dependence between more than two components. Interpretation of the models' features is given by looking at examples of the angular density on the trivariate simplex. The extremal dependence between five air pollutants is estimated to assess the air quality in the city of Leeds, UK. Models are fitted using the approximate likelihood and Bayesian approximate likelihood approaches. Previously only models whose mass is concentrated in the interior of the simplex were used for approximate likelihood. We show that if the model does put mass on subspaces of the simplex (edges, vertices, ...) then these have to be taken into account in the estimation procedure. Finally, the `R` package `ExtremalDep` (see Beranger et al. (2015)) was created and includes angular densities for all parametric models, a function to represent the density on the trivariate simplex, approximate likelihood and approximate Bayesian estimation methods are also available. This chapter corresponds to the accepted manuscript Beranger and Padoan (2015) to appear in the book *Extreme Value Modelling and Risks Analysis: Methods and Applications.* Eds. D. Dey and J. Yan. Chapman & Hall/CRC Press. The real data analysis and graphical examples are all reproducible using `ExtremalDep`.

Chapter 2 reviews some important facts on the extremes of stochastic processes defined as pointwise maxima of replicates of i.i.d. copies of a spatial or temporal process. Instead of taking the componentwise maxima of some random variables, stochastic processes are used and their convergence in the space of continuous functions is established. The limiting process is called max-stable process. We review the theory of max-stable processes and show its close relationship with Chapter 1. The canonical representations of such processes are defined and the max-stables models that arise from them are presented. Most often it is convenient to also

assume a parametric model for the dependence function; the characteristics of these models as well as their advantages and drawbacks are discussed for different correlation functions. These correlation functions do have an effect on the extremal dependence structure between spatial locations but their limitations are also noticeable. Furthermore we provide the finite-dimensional distribution for most max-stable models.

As mentioned above, the finite-dimensional distributions of some max-stable processes are not always available. Therefore, new methods are developed to approximate the likelihood. Buishand et al. (2008); Padoan et al. (2010); Blanchet and Davison (2011) among others, perform real data analyses for spatial environmental phenomena. They share the burden of approximating the full likelihood function in order to obtain a feasible model fitting, at a reasonable computational cost, when working with Brown-Resnick, Schlather and Smith marginal distributions. In Chapter 3 we give a particular attention to the work of Buishand et al. (2008) which considers an extension of the Brown-Resnick to the bivariate case, based on double sided Brownian motions. It also includes the idea of including a parameter measuring the amount of spatial dependence as suggested by de Haan and Pereira (2006). The bivariate marginal distributions of these types of processes were derived in de Haan and Zhou (2008). Our main contribution is the derivation of the analytical form of the marginal distributions for both univariate and bivariate Brown-Resnick processes. This allows different scenarios to estimate the spatial dependence parameter. However the results contained in this chapter haven't been published. After completing this work we discovered that identical results were derived and published in parallel by Huser and Davison (2013) for the generalized Brown-Resnick model. Our results are proved through an induction argument and a simulation study certifies their correctness. Results for univariate processes can also be recovered as a special case of Huser and Davison (2013).

The max-stable processes given in Chapter 2 have some common features: they all are strictly stationarity processes, i.e. a translation in space doesn't affect the distribution, and they are constructed from symmetric processes. It is however well known that environmental phenomena often exhibit skewed distributions and are not stationary, see for example Genton (2004); Arellano-Valle and Genton (2010) and Zhu et al. (2014). For the currently available models, only anisotropy can be included in the model through the correlation function. Therefore it is important to be able to study broader families of processes that allows for non-stationarity distributions and an asymmetric relationship among variables observed at different time or spatial lags. Chapter 4 provides the main contribution of this thesis proposing a new family of processes that overcome these challenges. Skew-symmetric distributions have the ability to describe a wide range of dependence behaviours and we provide an additive stochastic representation of Skew-Normal processes which can be shown to be non-stationary. It is then easy to define positive definite, non-stationary isotropic or anisotropic correlation functions and we derive a new family of max-stable processes – the extremal Skew-t process. The flexibility of the model is emphasized by visually showing the impact of the correlation function on the extremal dependence structure. Additionally, we establish the analytical form of its d-dimensional marginal distributions. The latter depends on a new class of asymmetric distribution: the non-central extended Skew-t distribution which includes the standard extended Skew-t distribution defined

by Arellano-Valle and Genton (2010). Density, distribution function and other properties such as marginal and conditional distributions are provided. Constructions of the model are obtained by considering the limiting distribution of appropriately rescaled Skew-t random variables but also through its spectral representation, used for random generation. Furthermore the joint tail dependence model given by Ledford and Tawn (1996) is used to show asymptotic independence. Finally, the extremal Skew-t model contains most of the popular models listed in Chapter 2 like the Extremal-t, Schlather and Smith models. The great flexibility and better fit provided by the extremal Skew-t model are proved through an application to extreme wind speeds in Oklahoma, USA. This chapter corresponds to the arXiv manuscript Beranger et al. (2015), that is under revision for the Scandinavian Journal of Statistics.

Univariate and multivariate extreme values are located in the tails of the distribution. This motivates us to investigate the estimation of tail densities in Chapter 5. The central role played by parametric models in extreme value analysis is illustrated in Chapter 1. Non-parametric estimators are often used to measure the accuracy of parametric models and for model selection. Chapter 5 proposes an innovative non-parametric method to analyze tail distributions, based on a modification of classical kernel estimators allowing a direct focus on the tail of the density. This method allows us to take the advantage of the good visualisation property of kernel estimators and their easy application to multivariate analysis. Moreover the asymptotic properties of our estimator are of the same order as usual kernel estimators. We show by simulation that our kernel based estimator is a good surrogate for the true underlying density and it can be used as reference for model selection. Efficiency improvement in model selection due to the kernel based tail density estimator is proved by comparison with histogram based estimators as considered in Perkins et al. (2013). An application to extreme temperatures in Sydney, Australia is provided with the aim is to project future extreme events which might severely affect humans and natural ecosystems. Geophysical models are available to simulate weather and climate as it would happen in the real world, and are compared to the observed climate. First the emphasis is put on climate models that render the extremes of maximum temperatures the most accurately. A second step consists of considering simultaneously large minimum and maximum values. This paper is in preparation for submission.

The final chapter is a summary and discussion on the results given in this thesis and links them to future research ideas.

Chapter 1

Extreme Dependence Models

Abstract

Extreme values of real phenomena are events that occur with low frequency, but can have a large impact on real life. These are, in many practical problems, high-dimensional by nature (e.g. Tawn, 1990; Coles and Tawn, 1991). To study these events is of fundamental importance. For this purpose, probabilistic models and statistical methods are in high demand. There are several approaches to modelling multivariate extremes as described in Falk et al. (2011), linked to some extent. We describe an approach for deriving multivariate extreme value models and we illustrate the main features of some flexible extremal dependence models. We compare them by showing their utility with a real data application, in particular analyzing the extremal dependence among several pollutants recorded in the city of Leeds, UK.

1.1 Introduction

Statistical analyses of extreme events are of crucial importance for risk assessment in many areas such as the financial market, telecommunications, industry, environment and health. For example governments and insurance companies need to statistically quantify the frequency of natural disasters in order to plan risk management and take preventive actions.

Several examples of univariate analysis are available, for instance in Coles (2001). Two main approaches are used in applications, the block-maximum and the peak over a threshold. These are based on the generalized extreme value (GEV) distribution and the generalized Pareto distridution (GPD), which are milestones of the extreme value theory, see e.g. Coles (2001, Ch. 3–4) and the references therein.

Many practical problems in finance, the environment, etc. are high-dimensional by nature, for example when analyzing the air quality in an area, the amount of pollution depends on the levels of different pollutants and the interaction between them. Today the extreme value theory provides a sufficiently mature framework for applications in the multivariate case. Indeed a large number of theoretical results and statistical methods and models are available, see for instance the monographs Resnick (2007), de Haan and Ferreira (2006), Falk et al. (2011), Beirlant et al.

(2006), Coles (2001) and Kotz and Nadarajah (2000). In this article we review some basic theoretical results on the extreme values of multivariate variables (multivariate extremes for brevity). With the block-maximum approach we explain what type of dependence structures can be described. We discuss the main features of some families of parametric extremal dependence models. By means of real data analysis we show the utility of these extremal dependence models when assessing the dependence of multivariate extremes. Their utility is also illustrated when estimating the probabilities that multivariate extreme events occur.

The analysis of real phenomena such as heavy rainfall, heat waves and so on is a challenging task. The first difficulty is the complexity of the data, i.e. observations are collected over space and time. In this case, theory deals with extremes of temporal- or spatial-processes (e.g. de Haan and Ferreira, 2006, Ch. 9). Examples of such statistical analysis are Davison et al. (2012), Davison and Gholamrezaee (2012), for a simple review see Padoan (2013a). This theory is closely linked to that of multivariate extremes presented here. The second difficulty is that the dependence of multivariate extremes is not always well captured by the models illustrated here. Ledford and Tawn (1996, 1997) have shown that in some applications a more suitable dependence structure is described by the so called *asymptotic independence*. This framework has been recently extended to continuous processes (e.g. De Haan and Zhou, 2011; Wadsworth and Tawn, 2012; Padoan, 2013c). These motivations make the multivariate extreme value theory a very active research field at present.

The paper is organized as follows. In Section 1.2 a definition of multivariate extremes is provided and the main characteristics are presented. In Section 1.3 some of the most popular extremal dependence models are described. In Section 1.4 some estimation methods are discussed and in Section 1.5 the analysis of the extremes of multiple pollutants is performed.

1.2 Multivariate Extremes

Applying the block-maximum approach to every component of a multivariate random vector gives rise to a definition of multivariate extremes. Specifically, for $d \in \mathbb{N}$, let $I = \{1, \dots, d\}$ be an index set and $\boldsymbol{X} = (X_1, \dots, X_d)$ be an $\mathbb{R}^d$-valued random vector with joint (probability) distribution function F and marginal distribution functions $F_j = F(\infty, \dots, x_j, \dots, \infty)$, $j \in I$. Suppose that $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are n independent and identically distributed (i.i.d.) copies of $\boldsymbol{X}$. The sample vector of componentwise maxima (sample maxima for brevity) is $\boldsymbol{M}_n = (M_{n,1}, \dots, M_{n,d})$, where $M_{n,j} = \max(X_{1,j}, \dots, X_{n,j})$.

Typically, in applications the distribution F is unknown and so the distribution of the sample maxima is also unknown. A possible solution is to study the asymptotic distribution of M_n as $n \to \infty$ and to use it as an approximation for a large but finite sample size, resulting in an approximate distribution for multivariate extremes. At a first glance, this notion of multivariate extremes may seem too simple to provide a useful approach for applications. However, a number of theoretical results justify its practical use. For example, with this definition of multivariate extremes, the dependence that arises is linked to the dependence that all the components of $\boldsymbol{X}$ are simultaneously large. Thus, by estimating these dependence structures we are also able to

estimate the probabilities that multiple exceedances occur.

1.2.1 Multivariate extreme value distributions

The asymptotic distribution of $\boldsymbol{M}_n$ is derived with a similar approach to the univariate case. Assume there are sequences of normalizing constants $\boldsymbol{a}_n = (a_{n1}, \ldots, a_{nd}) > \boldsymbol{0}$, with $\boldsymbol{0} = (0, \ldots, 0)$, and $\boldsymbol{b}_n = (b_{n1}, \ldots, b_{nd}) \in \mathbb{R}^d$ such that

$$\mathrm{pr}\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x}\right) = F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}), \qquad n \to \infty, \tag{1.1}$$

for all the continuity points $\boldsymbol{x}$ of a non-degenerate distribution G. The class of the limiting distributions in (1.1) is called *multivariate extreme value distributions* (MEVDs) (Resnick, 2007, p. 263). A distribution function F that satisfies the convergence result (1.1) is said to be in the *(maximum) domain of attraction* of G (de Haan and Ferreira, 2006, pp. 226–229). An attractive property of MEVDs is the *max-stability*. A distribution G on $\mathbb{R}^d$ is max-stable if for every $n \in \mathbb{N}$, there exists sequences $\boldsymbol{a}_n > \boldsymbol{0}$ and $\boldsymbol{b}_n \in \mathbb{R}^d$ such that

$$G(\boldsymbol{a}_n \, \boldsymbol{x} + \boldsymbol{b}_n) = G^{1/n}(\boldsymbol{x}), \tag{1.2}$$

(Resnick, 2007, Proposition 5.9). As a consequence, G is such that G^a is a distribution for every $a > 0$. A class of distributions that satisfies such a property is named *max-infinitely divisible* (max-id). More precisely, a distribution G on $\mathbb{R}^d$ is max-id, if for any $n \in \mathbb{N}$ there exists a distribution F_n such that $G = F_n^n$ (Resnick, 2007, p. 252). This means that G can always be defined through the distribution of the sample maxima of n i.i.d. random vectors.

In order to characterize the class of MEVDs we need to specify: a) the form of the marginal distributions, b) the form of the dependence structure.

a) To illustrate the first feature is fairly straightforward. If F converges, then so too does the marginal distributions F_j for all $j \in I$. Choosing a_{jn} and b_{jn} for all $j \in I$ as in de Haan and Ferreira (2006, Corollary 1.2.4), implies that each marginal distribution of G is a generalized extreme value (GEV), i.e.

$$G(\infty, \ldots, x_j, \ldots, \infty) = \exp\left[-\left\{1 + \xi_j\left(\frac{x_j - \mu_j}{\sigma_j}\right)\right\}_+^{-1/\xi_j}\right], \; j \in I,$$

where $(x)_+ = \max(0, x)$, $-\infty < \mu_j, \xi_j < \infty$, $\sigma_j > 0$ (de Haan and Ferreira, 2006, pp. 208–211). Because the marginal distributions are continuous then G is also continuous.

b) The explanation of the dependence form is more elaborate, although it is not complicated. The explanation is based on three steps: 1) G is transformed so that its marginal distributions are equal, 2) a Poisson point process (PPP) is used to represent the standardised distribution, 3) the dependence form is made explicit by means of a change of coordinates. Here are the steps.

1) Let $U_j(a) = F_j^{\leftarrow}(1 - 1/a)$, with $a > 1$, be the left-continuous inverse of F_j, for all $j \in I$.

The sequences a_{nj} and b_{nj} in (1.1) are such that for all $y_j > 0$,

$$\lim_{n\to\infty} \frac{U_j(ny_j) - b_n}{a_n} = \frac{\sigma_j(y_j^{\xi_j} - 1)}{\xi_j} + \mu_j, \quad j \in I,$$

and therefore

$$\lim_{n\to\infty} F^n\{U_1(ny_1), \ldots, U_d(ny_d)\} = G\left(\frac{\sigma_1(y_1^{\xi_1} - 1)}{\xi_1} + \mu_1, \ldots, \frac{\sigma_d(y_d^{\xi_d} - 1)}{\xi_d} + \mu_d\right) \equiv G_0(\boldsymbol{y}), \tag{1.3}$$

for all continuity points $\boldsymbol{y} > \mathbf{0}$ of G_0 (see de Haan and Ferreira, 2006, Theorems 1.1.6, 6.1.1). G_0 is a MEVD with identical unit Fréchet marginal distributions.

Now, for all $\boldsymbol{y} > \mathbf{0}$ such that $0 < G_0(\boldsymbol{y}) < 1$, by taking the logarithm on the right and left side of (1.3) and using a first order Taylor expansion of $\log F\{U_1(ny_1), \ldots, U_d(ny_d)\}$, as $n \to \infty$, it follows that

$$\lim_{n\to\infty} n[1 - F\{U_1(ny_1), \ldots, U_d(ny_d)\}] = -\log G_0(\boldsymbol{y}) \equiv V(\boldsymbol{y}). \tag{1.4}$$

The function V, named *exponent (dependence) function*, represents the dependence structure of multiple extremes (extremal dependence for brevity). According to (1.4) the derivation of V depends on the functional form of F. In most of the practical problems the latter is unknown. A possible solution is obtained exploiting the max-id property of G_0, which says that every max-id distribution permits a PPP representation, see Resnick (2007, pp. 257–262) and Falk et al. (2011, pp. 141–142).

2) Let $N_n(\cdot)$ be a PPP defined by

$$N_n(\mathcal{A}) := \sum_{i=1}^{\infty} \mathbb{1}_{\{\boldsymbol{P}_i\}}(\mathcal{A}), \quad \mathbb{1}_{\{\boldsymbol{P}_i\}}(\mathcal{A}) = \begin{cases} 1, & \boldsymbol{P}_i \in \mathcal{A}, \\ 0, & \boldsymbol{P}_i \notin \mathcal{A}, \end{cases}$$

where $\mathcal{A} \subset \mathbb{A}$ with $\mathbb{A} := (0, \infty) \times \mathbb{R}_+^d$,

$$\boldsymbol{P}_i = \left[\frac{i}{n}, \left\{1 + \xi_1\left(\frac{X_{i1} - b_{n1}}{a_{n1}}\right)\right\}^{\frac{1}{\xi_1}}, \ldots, \left\{1 + \xi_1\left(\frac{X_{id} - b_{nd}}{a_{nd}}\right)\right\}^{\frac{1}{\xi_d}}\right],$$

for every $n \in \mathbb{N}$ and $\boldsymbol{X}_i$, $i = 1, 2, \ldots$ are i.i.d random vectors with distribution F. The intensity measure is $\zeta \times \nu_n$ where ζ is the Lebesgue measure and for every $n \in \mathbb{N}$ and all critical regions defined by $\mathcal{B}_{\boldsymbol{y}} := \mathbb{R}_+^d \setminus [\mathbf{0}, \boldsymbol{y}]$ with $\boldsymbol{y} > 0$,

$$\nu_n(\mathcal{B}_{\boldsymbol{y}}) = n[1 - F\{U_1(ny_1), \ldots, U_d(ny_d)\}],$$

is a finite measure. If the limit in (1.3) holds, then N_n converges weakly to N as $n \to \infty$, i.e. a

PPP with intensity measure $\zeta \times \nu$ where

$$\nu(\mathcal{B}_{\boldsymbol{y}}) = \nu\{(\boldsymbol{v} \in \mathbb{R}_+^d : v_1 > y \,\text{or}\, \ldots \,\text{or}\, v_d > y_d)\} \equiv V(\boldsymbol{y}), \quad \boldsymbol{y} > 0,$$

is a finite measure, named *exponent measure* (see de Haan and Ferreira, 2006, Theorems 6.1.5, 6.1.11). See Appendix A for details on the convergence of simple point processes to PPP. Observe that ν must concentrate on $\overline{\mathbb{R}} = \mathbb{R}_+^d \backslash \{\mathbf{0}\}$ in order to be uniquely determined. Also, ν must satisfy $\nu(\boldsymbol{\infty}) = 0$, see Falk et al. (2011, p. 143) for details.

This essentially means that numbering the rescaled observations that fall in a critical region, e.g. see the shaded sets in the left panels of Figure 1.1, where at least one coordinate is large, makes it possible for (1.3) to be computed using the void probability of N, that is

$$\begin{aligned} G_0(\boldsymbol{y}) &= \text{pr}[N\{(0,1] \times \mathcal{B}_{\boldsymbol{y}}\} = 0] \\ &= \exp(-[\zeta\{(0,1]\} \times \nu(\mathcal{B}_{\boldsymbol{y}})]) \\ &= \exp\{-V(\boldsymbol{y})\} \quad \boldsymbol{y} > \mathbf{0}. \end{aligned} \tag{1.5}$$

From Figure 1.1 we see that in the case of strong dependence (top-left panel) all the coordinates of the extremes are large, while in the case of weak dependence (bottom-left panels) only one coordinate of the extremes is large.

At this time it remains to specify the structure of the exponent measure. This task is simpler to fulfill when working with pseudo-polar coordinates.

3) With unit Fréchet margins, the stability property (1.2) can be rephrased by $G_0^a(a\boldsymbol{y}) = G_0(\boldsymbol{y})$ for any $a > 0$, implying that ν satisfies the homogeneity property

$$\nu(a\mathcal{B}_{\boldsymbol{y}}) = \nu(\mathcal{B}_{\boldsymbol{y}})/a, \tag{1.6}$$

for all $\mathcal{B}_{\boldsymbol{y}} \subset \overline{\mathbb{R}}$, where $\mathcal{B}_{\boldsymbol{y}} := \overline{\mathbb{R}} \backslash (\mathbf{0}, \boldsymbol{y}]$ with $\boldsymbol{y} > \mathbf{0}$. Note that for a Borel set $\mathcal{B} \subset \overline{\mathbb{R}}$ we have $a\mathcal{B} = \{a\boldsymbol{v} : \boldsymbol{v} \in \mathcal{B}\}$ and $\mathcal{B}_{a\boldsymbol{y}} = a\mathcal{B}_{\boldsymbol{y}}$. Now, let

$$\mathbb{W} := (\boldsymbol{v} \in \overline{\mathbb{R}} : v_1 + \ldots + v_d = 1),$$

be the unit simplex on $\overline{\mathbb{R}}$ (simplex for brevity), where $d-1$ variables are free to vary and one is fixed, e.g. $v_d = 1 - (v_1 + \cdots + v_{d-1})$. For any $\boldsymbol{v} \in \mathbb{R}_+^d$, with the sum-norm, $\|\boldsymbol{v}\| = |v_1| + \cdots + |v_d|$, we measure the distance of $\boldsymbol{v}$ from $\mathbf{0}$. Other norms can also be considered (e.g. Resnick, 2007, pp. 270–274). We consider the one-to-one transformation $Q : \overline{\mathbb{R}} \to (0, \infty) \times \mathbb{W}$, given by

$$(r, \boldsymbol{w}) := Q(\boldsymbol{v}) = (\|\boldsymbol{v}\|, \|\boldsymbol{v}\|^{-1}\boldsymbol{v}), \quad \boldsymbol{v} \in \overline{\mathbb{R}}. \tag{1.7}$$

By means of this, the induced measure is $\psi := \nu * Q$, i.e. $\psi(\mathcal{W}_r) = \nu\{Q^{\leftarrow}(\mathcal{W}_r)\}$ for all sets

$\mathcal{W}_r = r \times \mathcal{W}$ with $r > 0$ and $\mathcal{W} \subset \mathbb{W}$, is generated. Then, from the property (1.6) it follows that

$$\begin{aligned}\psi(\mathcal{W}_r) &= \nu\{(\boldsymbol{v} \in \overline{\mathbb{R}} : \|\boldsymbol{v}\| > r, \boldsymbol{v}/\|\boldsymbol{v}\| \in \mathcal{W})\} \\ &= \nu\{(r\boldsymbol{u} \in \overline{\mathbb{R}} : \|\boldsymbol{u}\| > 1, \boldsymbol{u}/\|\boldsymbol{u}\| \in \mathcal{W})\} \\ &= r^{-1}H'(\mathcal{W}),\end{aligned}$$

where $H'(\mathcal{W}) := \nu\{(\boldsymbol{u} \in \overline{\mathbb{R}} : \|\boldsymbol{u}\| > 1, \boldsymbol{u}/\|\boldsymbol{u}\| \in \mathcal{W})\}$. The benefit of transforming the coordinates into pseudo-polar is that the measure ν becomes a product of two independent measures: the *radial measure* $(1/r)$ and *spectral measure* or *angular measure* (H') (e.g. Falk et al., 2011, p. 145). The first measures the intensity (or distance) of the points from the origin and the second measures the angular spread (or direction) of the points. This result is known as the *spectral decomposition* (de Haan and Resnick, 1977). Hereafter we will use the term angular measure.

The density of ψ is $\mathrm{d}\psi(r, \boldsymbol{w}) = r^{-2}\mathrm{d}r \times \mathrm{d}H'(\boldsymbol{w})$ for all $r > 0$ and $\boldsymbol{w} \in \mathbb{W}$, by means of which we obtain the explicit form

$$\begin{aligned}\nu(\mathcal{B}_{\boldsymbol{y}}) &= \psi\{Q(\boldsymbol{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ or} \ldots \text{or } v_d > y_d)\} \\ &= \psi[\{(r, \boldsymbol{w}) \in (0, \infty) \times \mathbb{W} : r > \min(y_j/w_j, j \in I)\}] \\ &= \int_{\mathbb{W}} \int_{\min(y_j/w_j, j \in I)}^{\infty} r^{-2}\mathrm{d}r\mathrm{d}H'(\boldsymbol{w}) \\ &= \int_{\mathbb{W}} \max_{j \in I} (w_j/y_j)\, \mathrm{d}H'(\boldsymbol{w}).\end{aligned} \tag{1.8}$$

In pseudo-polar coordinates, extremes are the values whose radial component is higher than a high threshold, see the red points in the middle panels of Figure 1.1. The angular components are concentrated around the center of the simplex, in the case of strong dependence (middle-top panel), while they are concentrated around the vertices of the simplex (middle-bottom panel), in the case of weak dependence.

The measure H' can be any finite measure on $\mathbb{W}$ satisfying the first moment conditions

$$\int_{\mathbb{W}} w_j\, \mathrm{d}H'(\boldsymbol{w}) = 1, \quad \forall\, j \in I.$$

This guarantees that the marginal distributions of G_0 are unit Fréchet. If H' satisfies the first moment conditions, then the total mass is equal to

$$H'(\mathbb{W}) = \int_{\mathbb{W}} (w_1 + \cdots + w_d)\mathrm{d}H'(\boldsymbol{w}) = \sum_{j \in I} \int_{\mathbb{W}} w_j \mathrm{d}H'(\boldsymbol{w}) = d.$$

So setting $H := H'/H'(\mathbb{W})$, then H is a probability measure satisfying

$$\int_{\mathbb{W}} w_j \mathrm{d}H(\boldsymbol{w}) = 1/d, \quad \forall\, j \in I. \tag{1.9}$$

Concluding, combining (1.3), (1.4), (1.5) and (1.8) all together, we have that a MEVD with unit

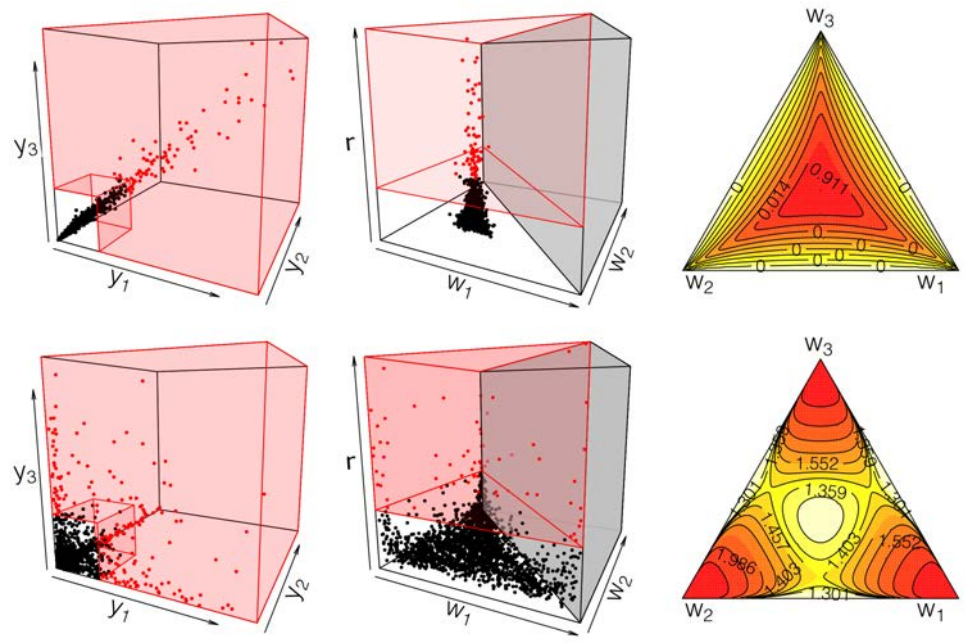

Figure 1.1: Examples of critical regions in $\mathbb{R}^3_+$ (left-panels) and its representation in pseudo-polar coordinates (middle-panels). Red points are the extremes with strong (top-panels) and weak (bottom-panels) dependence. Right panels display the angular densities on the simplex.

Fréchet margins is equal to

$$G_0(\boldsymbol{y}) = \exp\left\{-d\int_{\mathbb{W}} \max_{j\in I}(w_j/y_j)\,\mathrm{d}H(\boldsymbol{w})\right\}. \tag{1.10}$$

1.2.2 Angular densities

The measure H can place mass on the interior as well as on other subspaces of the simplex, such as the edges and the vertices. Thus H can have several densities that lie on these sets, which are named *angular densities*. Coles and Tawn (1991) described a way to derive the angular densities when G is absolutely continuous (see also Resnick, 2007, Example. 5.13).

Specifically, let $S := \mathbb{P}(I)\backslash\varnothing$, where $\mathbb{P}(I)$ is the power set of I and $\mathcal{S}$ be the index set that takes values in S. Given fixed d, the sets

$$\mathbb{W}_{d,\mathcal{S}} = (\boldsymbol{w}\in\mathbb{W} : w_j = 0,\ \text{if } j\notin\mathcal{S};\ w_j > 0 \text{ if } j\in\mathcal{S}),$$

for all $\mathcal{S}\in S$ provide a partition of $\mathbb{W}$ in 2^d-1 subsets. Similar to the simplex, there are $k-1$ variables w_j in $\mathbb{W}_{d,\mathcal{S}}$ that are free to vary, where $j\in\mathcal{S}$ and $k = |\mathcal{S}|$ denotes the size of $\mathcal{S}$. We denote by $h_{d,\mathcal{S}}$ the density that lies on the subspace $\mathbb{W}_{d,\mathcal{S}}$, where $\mathcal{S}\in S$. When the latter is a vertex $\boldsymbol{e}_j$ of the simplex $\mathbb{W}$, for any $j\in I$, then the density is a point mass, that is $h_{d,\mathcal{S}} = H(\{\boldsymbol{e}_j\})$.

Let $\mathcal{S} = \{i_1,\ldots,i_k\}\subset I$, when G_0 is absolutely continuous the angular density for any

$\boldsymbol{y} \in \mathbb{R}_+^d$ is

$$h_{d,\mathcal{S}}\left(\frac{y_{i_1}}{\sum_{i\in\mathcal{S}} y_i}, \cdots, \frac{y_{i_{k-1}}}{\sum_{i\in\mathcal{S}} y_i}\right) = -\left(\sum_{i\in\mathcal{S}} y_i\right)^{(k+1)} \lim_{\substack{y_j\to 0,\\ j\notin\mathcal{S}}} \frac{\partial^k V}{\partial y_{i_1}\cdots\partial y_{i_k}}(\boldsymbol{y}). \tag{1.11}$$

Two examples of a tridimensional angular density in the interior of the simplex are reported in the right panels of Figure 1.1. These are the densities of a symmetric logistic model (Gumbel, 1960) with a strong and weak dependence. When $\mathcal{S} = \{i\}$ for any $i \in I$ the angular density $h_{d,\mathcal{S}}$ represents the mass of H at the vertex $\boldsymbol{e}_j$ with $j = i$, thus (1.11) reduces into

$$h_{d,\mathcal{S}} = H(\{\boldsymbol{e}_i\}) = -y_i^{(2)} \lim_{y_j\to 0, j\notin\mathcal{S}} \frac{\partial V}{\partial y_i}(\boldsymbol{y}). \tag{1.12}$$

In the bivariate case these results are equal to the ones obtained by Pickands (1981). Kotz and Nadarajah (2000) discussed the bivariate case in the following terms. With $d = 2$ the unit simplex $\mathbb{W} = [0, 1]$ can be partitioned into

$$\mathbb{W}_{2,\{1\}} = \{(1, 0)\}, \quad \mathbb{W}_{2,\{2\}} = \{(0, 1)\}, \quad \mathbb{W}_{2,\{1,2\}} = \{(w, 1 - w), w \in (0, 1)\}.$$

The densities that lie on them are

$$h_{2,\{1\}} = H(\{0\}) = -y_1^2 \lim_{y_2\to 0} \frac{\partial V}{\partial y_1}(y_1, y_2),$$

$$h_{2,\{2\}} = H(\{1\}) = -y_2^2 \lim_{y_1\to 0} \frac{\partial V}{\partial y_2}(y_1, y_2),$$

and

$$h_{2,\{1,2\}}(w) = -\frac{\partial^2 V}{\partial y_1\,\partial y_2}(w, 1 - w).$$

respectively, for any $y_1, y_2 > 0$. The first two densities describe the case when extremes are only observed in one variable. While the third density describes the case when extremes are observed in both variables.

1.2.3 Extremal dependence

From (1.5) it emerges that the extremal dependence is expressed through the exponent function. This is a map from $\mathbb{R}_+^d$ to $(0, \infty)$ satisfying the properties:

1. is a continuous function and homogeneous of order -1, the latter meaning that $V(a\boldsymbol{y}) = a^{-1}V(\boldsymbol{y})$ for all $a > 0$;
2. is a convex function, that is $V(a\boldsymbol{y} + (1 - a)\boldsymbol{y}') \leq aV(\boldsymbol{y}) + (1 - a)V(\boldsymbol{y}')$, for $a \in [0, 1]$ and $\boldsymbol{y}, \boldsymbol{y}' \in \mathbb{R}_+^d$;
3. $\max(1/y_1, \ldots, 1/y_d) \leq V(\boldsymbol{y}) \leq (1/y_1 + \ldots + 1/y_d)$, with the lower and upper limits representing the complete dependence and independence cases respectively.

See de Haan and Ferreira (2006, pp. 223–226) for details. In summary, let $\boldsymbol{Y}$ be a random vector with distribution (1.10). When H places the total mass 1 on the center of the simplex $(1/d, \ldots, 1/d)$, then $Y_1 = Y_2 = \cdots = Y_d$ almost surely and hence $G_0(\boldsymbol{y}) = \exp\{\max(1/y_1, \ldots, 1/y_d)\}$. When H places mass $1/d$ on $\boldsymbol{e}_j$ for all $j \in I$, i.e. the vertices of the simplex, then $Y_1, \ldots, Y_d$ are independent and hence $G_0(\boldsymbol{y}) = \exp(1/y_1 + \ldots + 1/y_d)$. This rephrased for a random vector $\boldsymbol{X}$ with distribution (1.1) becomes

$$\min\{G_1(x_1), \ldots, G_d(x_d)\} \leq G(\boldsymbol{x}) \leq G_1(x_1) \cdot \ldots \cdot G_d(x_d), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

In order to visualise the exponent function more easily, its restriction in the simplex is usually considered. This is a function $A : \mathbb{W} \to [1/d, 1]$, named the *Pickands dependence* function (Pickands, 1981), defined by

$$A(\boldsymbol{t}) := d \int_{\mathbb{W}} \max_{j \in I} (w_j \, t_j) \, \mathrm{d}H(\boldsymbol{w}),$$

where $z_j = 1/y_j$, $j \in I$, $t_j = z_j/(z_1 + \cdots + z_d)$ with $j = 1, \ldots, d-1$ and $t_d = 1 - (t_1 + \cdots + t_{d-1})$. A inherits the above properties from V with the obvious modifications. In particular, $1/d \leq \max(t_1, \ldots, t_d) \leq A(\boldsymbol{t}) \leq 1$, where lower and upper bounds represent the complete dependence and independence cases, and for the homogeneity property of A the exponent function can be rewritten as

$$V(\boldsymbol{z}) = (z_1 + \cdots + z_d) A(t_1, \ldots, t_d), \quad \boldsymbol{z} \in \mathbb{R}_+^d.$$

The exponential function can be profitably used in several ways. First, an important summary of the extremal dependence is given by

$$\vartheta = V(1, \ldots, 1) = d \int_{\mathbb{W}} \max_{j \in I}(w_j) \mathrm{d}H(\boldsymbol{w}). \tag{1.13}$$

This is named the *extremal coefficient* (Smith, 1990) and it represents the (fractional) number of independent components of the random vector $\boldsymbol{Y}$. The coefficient takes values in $[1, d]$, depending on whether the measure H concentrates near the center or the vertices of the simplex. The bounds regard the cases of complete dependence and independence.

Second, for any $\boldsymbol{y} > \boldsymbol{0}$ and failure region

$$\mathcal{F}_{\boldsymbol{y}} = (\boldsymbol{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ and} \ldots \text{and } v_d > y_d), \tag{1.14}$$

the *tail dependence* function (Nikoloulopoulos et al., 2009; de Haan and Ferreira, 2006, p. 225) is defined by

$$R(\boldsymbol{y}) := \nu\{(\boldsymbol{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ and} \ldots \text{and } v_d > y_d)\} \equiv \nu(\mathcal{F}_{\boldsymbol{y}}), \quad \boldsymbol{y} > \boldsymbol{0}.$$

This counts the number of observations that fall in the failure region, i.e. all their coordinates are simultaneously large. The tail dependence function is related to the exponent function by the inclusion-exclusion principle. Using similar arguments to those in (1.8) and (1.9) it follows

that

$$R(\boldsymbol{y}) = d \int_{\mathbb{W}} \min_{j \in I}(w_j / y_j) \mathrm{d}H(\boldsymbol{w}) \quad \boldsymbol{y} > \boldsymbol{0}. \tag{1.15}$$

By means of the tail dependence function, another important summary of the dependence between the components of $\boldsymbol{Y}$ is obtained. The *coefficient of upper tail dependence* is given by

$$\chi = R(1, \ldots, 1) = d \int_{\mathbb{W}} \min_{j \in I}(w_j) \mathrm{d}H(\boldsymbol{w}). \tag{1.16}$$

It measures the strength of dependence in the tail of the distribution of $\boldsymbol{Y}$ or in other terms the probability that all the components of $\boldsymbol{Y}$ are simultaneously large. This coefficient was introduced in the bivariate case by Joe (1997, Ch. 2) and extended to the multivariate case by Li (2009). When H concentrates near the center or on the vertices of the simplex, then $\chi > 0$ or $\chi = 0$ respectively. In these cases we say that $\boldsymbol{Y}$ is upper tail dependent or independent.

In addition, the exponent and the tail dependence functions can be used for approximating the probability that certain types of extreme events will occur. Specifically, let $\boldsymbol{Y}$ be a random vector with unit Pareto margins. F is in the domain of attraction of a MEVD with Fréchet margins. From (1.4) and for the homogeneity property of V we have that $\{1 - F(n\boldsymbol{y})\} \approx V(n\boldsymbol{y})$ for large n. Then, for the relations (1.8) and (1.9), the approximating result follows

$$\mathrm{pr}(Y_1 > y_1 \text{ or } \ldots \text{ or } Y_d > y_d) \approx d \int_{\mathbb{W}} \max_{j \in I} (w_j / y_j) \, \mathrm{d}H(\boldsymbol{w}), \tag{1.17}$$

when $y_1, \ldots, y_d$ are high enough thresholds. Furthermore, with similar arguments to those in Section 1.2.1 we have that

$$\lim_{n \to \infty} n \bar{F}(ny_1, \ldots, ny_d) = R(\boldsymbol{y}),$$

where $\bar{F}$ is the survivor function of $\boldsymbol{Y}$. R has the same homogeneity property of V. Hence, $\bar{F}(n\boldsymbol{y}) \approx R(n\boldsymbol{y})$ for large n. Then, for the relation (1.15), the approximating result also follows

$$\mathrm{pr}(Y_1 > y_1 \text{ and} \ldots \text{and } Y_d > y_d) \approx d \int_{\mathbb{W}} \min_{j \in I} (w_j / y_j) \, \mathrm{d}H(\boldsymbol{w}), \tag{1.18}$$

when $y_1, \ldots, y_d$ are high enough thresholds.

Lastly, when $\chi = 0$ the elements of $\boldsymbol{Y}$ are independent in the limit. However, they may still be dependent for large but finite samples. Ledford and Tawn (1996) proposed another dependence measure in order to capture this feature. For brevity, we focus on the bivariate case. Suppose that $\bar{F}$ for $y \to \infty$ satisfies the condition

$$\bar{F}(y, y) \approx y^{-1/\tau} \mathscr{L}(y), \quad 0 < \tau \leq 1,$$

where $\mathscr{L}$ is a slowly function, i.e. $\mathscr{L}(ay)/\mathscr{L}(y) \to 1$ as $y \to \infty$ for any $a > 0$. Then for large y, assuming $\mathscr{L}$ constant, different tail behaviours are covered. The case $\chi > 0$ is reached when $\tau = 1$ and so the variables are asymptotically dependent. When $1/2 < \tau < 1$ this means that $\chi = 0$ and so the variables are asymptotically independent, but they are still positively associated and the value of τ expresses the degree (see Ledford and Tawn, 1996, for details).

1.3 Parametric models for the extremal dependence

From the previous sections, it emerges that both the exponent and tail dependence functions depend on the angular measure. There is no unique angular measure that generates the extremal dependence, any finite measure that satisfies the first moment conditions is suitable. In order to represent the extremal dependence, in principle it is insufficient to use a parametric family of models for the distribution function of the angular measure. However, flexible classes of parametric models can still be useful for applications, e.g. see Tawn (1990), Coles and Tawn (1991) and Boldi and Davison (2007) to name a few. To this end, in previous years different parametric extremal dependence models have been introduced in the literature. A fairly comprehensive overview can be found in Kotz and Nadarajah (2000, Section 3.4), Coles (2001, Section 8.2.1), Beirlant et al. (2006, Section 9.2.2) and Padoan (2013b). In the next sections we describe some of the most popular models.

1.3.1 Asymmetric logistic model

The multivariate asymmetric logistic model is an extension of the symmetric, introduced by Tawn (1990) (see also Coles and Tawn, 1991) for modelling extremes in complex environmental applications.

Let S and $\mathcal{S}$ as in Section 1.2.2 and $N_{\mathcal{S}}$ be a Poisson random variable with rate $1/\tau_{\mathcal{S}}$. This describes the number of storm events, $n_{\mathcal{S}}$, that takes place on the sites $\mathcal{S}$ in a time interval. Given $n_{\mathcal{S}}$, for any site $j \in \mathcal{S}$, let $\{X_{j,\mathcal{S};i}, i = 1, \dots, n_{\mathcal{S}}\}$ be a sequence of i.i.d. random variables that describe an environmental episode such as rain. For a fixed i, $\{X_{j,\mathcal{S};i}\}_{j\in\mathcal{S}}$ is assumed to be a dependent sequence. The maximum amount of rain observed at j is $X_{j,\mathcal{S}} = \max_{i=1\dots,n_{\mathcal{S}}}\{X_{j,\mathcal{S};i}\}$. Let $A_{\mathcal{S}}$ be a random effect with a positive stable distribution and stability parameter $\alpha_{\mathcal{S}} \geq 1$ (Nolan, 2003), representing an unrecorded additional piece of information on storm events. Assume $\{X_{j,\mathcal{S}}\}_{j\in\mathcal{S}}|\alpha_{\mathcal{S}}$ as an independent sequence. Define $Y_j = \max_{\mathcal{S}\in S_j}\{X_{j,\mathcal{S}}\}$, where $S_j \subset S$ contains all nonempty sets including j and so the maximum is over all the storm events involving j. Then, the exponent function of the joint survival function of $(Y_1, \dots, Y_d)$, after transforming the margins into unit exponential variables, is

$$V(\boldsymbol{y};\boldsymbol{\theta}) = \sum_{\mathcal{S}\in S}\Big\{\sum_{j\in\mathcal{S}}(\beta_{j,\mathcal{S}}y_j^{-1})^{\alpha_{\mathcal{S}}}\Big\}^{1/\alpha_{\mathcal{S}}}, \quad \boldsymbol{y} \in \mathbb{R}_+^d,$$

where $\boldsymbol{\theta} = \{\alpha_{\mathcal{S}}, \beta_{j,\mathcal{S}}\}_{\mathcal{S}\in S}$, $\alpha_{\mathcal{S}} \geq 1$, $\beta_{\mathcal{S}} = \tau_{\mathcal{S}}/\sum_{\mathcal{S}\in S_j}\tau_{\mathcal{S}}$ and $\beta_{j,\mathcal{S}} = 0$ if $j \notin \mathcal{S}$, and for $j \in I$, $0 \leq \beta_{j,\mathcal{S}} \leq 1$ and $\sum_{\mathcal{S}\in S}\beta_{j,\mathcal{S}} = 1$. The parameter $\beta_{j,\mathcal{S}}$ represents the probability that the maximum value observed at j is attributed to a storm event involving the sites of $\mathcal{S}$. The number of the model parameters is $2^{d-1}(d+2) - (2d+1)$.

In this case the angular measure places mass on all the subspaces of the simplex. From (1.11) it follows that the angular density is, for every $\mathcal{S} \in S$ and all $\boldsymbol{w} \in \mathbb{W}_{d,\mathcal{S}}$ equal to

$$h_{d,\mathcal{S}}(\boldsymbol{w};\boldsymbol{\theta}) = \prod_{i=1}^{k-1}(i\alpha_{\mathcal{S}} - 1)\prod_{j\in\mathcal{S}}\beta_{j,\mathcal{S}}^{\alpha_{\mathcal{S}}}w_j^{-(\alpha_{\mathcal{S}}+1)}\Big\{\sum_{j\in\mathcal{S}}(\beta_{j,\mathcal{S}}/w_j)^{\alpha_{\mathcal{S}}}\Big\}^{1/\alpha_{\mathcal{S}}-k}.$$

When $\mathcal{S} = I$, $\alpha_{\mathcal{S}} = \alpha$, $\beta_{j,\mathcal{S}} = \beta_j$ and so the angular density on the interior of the simplex simplifies to

$$h(\boldsymbol{w};\boldsymbol{\theta}) = \prod_{i=1}^{d-1}(i\alpha - 1)\prod_{j\in I}\beta_j^{\alpha} w_j^{-(\alpha+1)}\Big\{\sum_{j\in I}(\beta_j/w_j)^{\alpha}\Big\}^{1/\alpha-d}, \quad \boldsymbol{w}\in\mathbb{W}.$$

When $\mathcal{S} = \{j\}$, for all $j \in I$, then from (1.12) it follows that the point mass at each extreme point of the simplex is $h_{d,\mathcal{S}} = \beta_{j,s}$.

For example in the bivariate case, the conditions on the parameters are $\beta_{1,\{1\}} + \beta_{1,\{1,2\}} = 1$ and $\beta_{2,\{2\}} + \beta_{2,\{1,2\}} = 1$, so the masses at the corners of $S_2 = [0,1]$ are given by $h_{2,\{1\}} = 1 - \beta_1$ and $h_{2,\{2\}} = 1 - \beta_2$, where for simplicity $\beta_{1,\{1,2\}} = \beta_1$ and $\beta_{2,\{1,2\}} = \beta_2$, while the density in the interior of the simplex, for $0 < w < 1$, is

$$h_{2,\{1,2\}}(w) = (\alpha-1)(\beta_1\beta_2)^{\alpha}\{w(1-w)\}^{\alpha-2}[(\beta_1(1-w))^{\alpha} + (\beta_2 w)^{\alpha}]^{1/\alpha-2}.$$

The top row of Figure 1.2 illustrates some examples of trivariate angular densities for different values of the parameters $\boldsymbol{\theta} = (\alpha, \beta_1, \beta_2, \beta_3)$, where the subscript of the index set $\mathcal{S} = \{1,2,3\}$ has been omitted for simplicity. The values of the parameters are, from left to right $\{(5.75, 0.5, 0.5, 0.5); (1.01, 0.9, 0.9, 0.9); (1.25, 0.5, 0.5, 0.5); (1.4, 0.7, 0.15, 0.15)\}$. The first panel shows that with large values of α and equal values of the other parameters, the case of strong dependence among the variables is obtained. The mass is mainly concentrated towards the center of the simplex. The second panel shows that when α is close to 1 and the other parameters are equal, the case of weak dependence is obtained. The mass is concentrated on the vertices of the simplex. The third panel shows the case of a symmetric dependence structure with the mass near the corners of the simplex but not along the edges. Finally, the fourth panel shows a case of an asymmetric dependence structure where the mass tends to be closer to the components whose corresponding values of β are high.

1.3.2 Tilted Dirichlet model

Extremal dependence models with an angular measure that places mass on the interior, vertices and edges of the simplex are more flexible than those with a measure that concentrates only on the interior. An example is the asymmetric logistic model versus the symmetric. However, the former has too many parameters to estimate, so parsimonious models may be preferred. In order to derive a parametric model for the angular density whose mass concentrates on the interior of the simplex, Coles and Tawn (1991) proposed the following method. Consider a continuous function $h' : \mathbb{W} \to [0,\infty)$ such that $m_j = \int_{S_d} v_j\, h'(\boldsymbol{v})\mathrm{d}\boldsymbol{v} < \infty$ for all $j \in I$. Then, the function

$$h(\boldsymbol{w}) = d^{-1}(m_1 w_1 + \cdots + m_d w_d)^{-(d+1)} h'\{\boldsymbol{m}\boldsymbol{w}/(m_1 w_1 + \cdots + m_d w_d)\}, \boldsymbol{w}\in\mathbb{W}$$

is a valid angular density. It satisfies the first moment conditions (1.9) and its mass is centered at $(1/d, \ldots, 1/d)$ and integrates to one. For example, if h' is the density of the Dirichlet distribution,

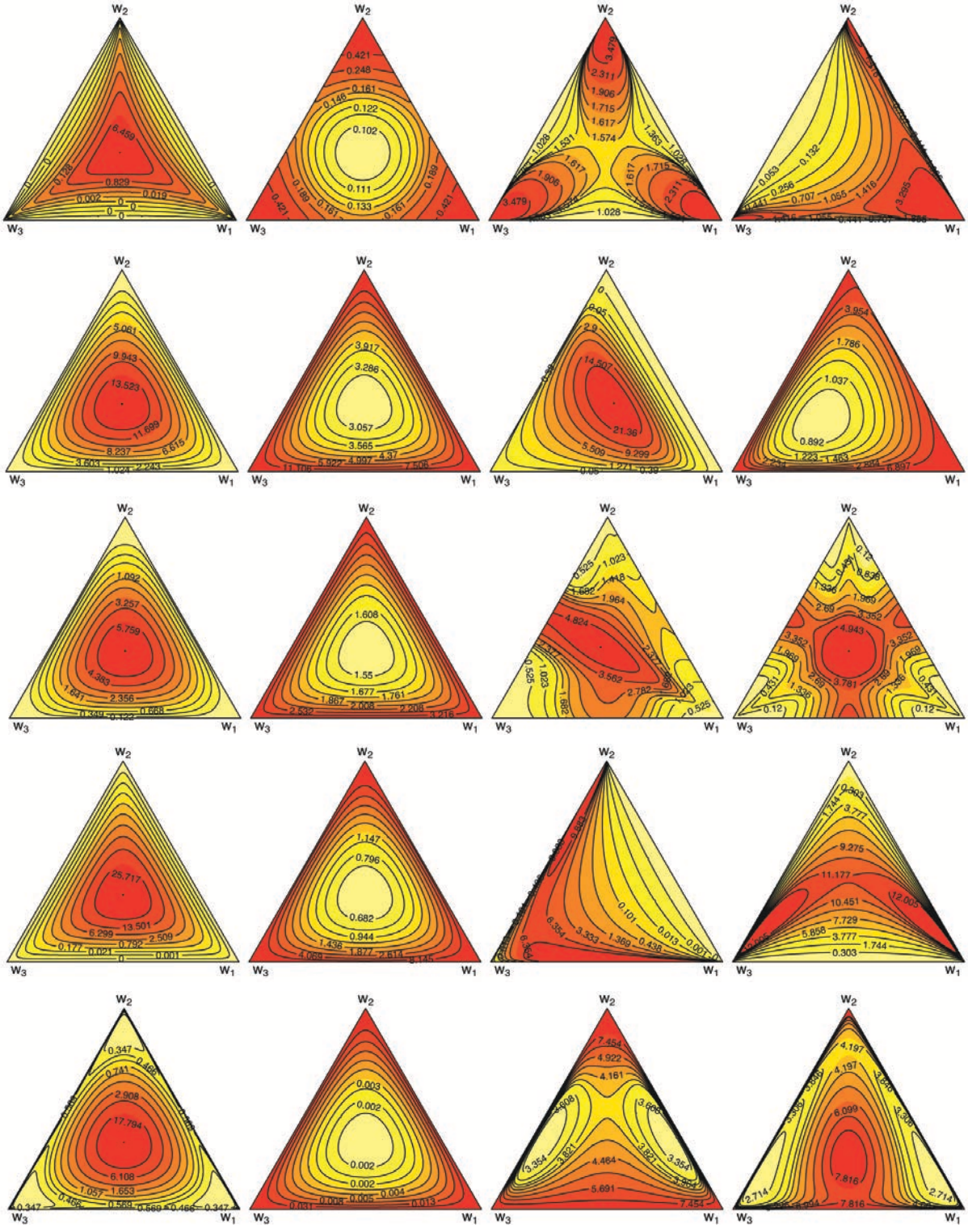

Figure 1.2: Examples of trivariate angular densities for the Asymmetric Logistic, Tilted Dirichlet, Pairwise Beta, Hüsler-Reiss and Extremal-t models from top to bottom.

then we obtain the angular density

$$h(\boldsymbol{w};\boldsymbol{\theta}) = \frac{\Gamma(\sum_{j\in I}\alpha_j + 1)}{d(\sum_{j\in I}\alpha_j w_j)^{d+1}} \prod_{j=1}^{d} \frac{\alpha_j}{\Gamma(\alpha_j)} \left(\frac{\alpha_j w_j}{\sum_{j\in I}\alpha_j w_j}\right)^{\alpha_j - 1}, \boldsymbol{w} \in \mathbb{W}, \tag{1.19}$$

where $\boldsymbol{\theta} = \{\alpha_j > 0\}_{j \in I}$. This density is asymmetric and it becomes symmetric when $\alpha_1 = \cdots = \alpha_d$. Extremes are independent or completely dependent when for all $j \in I$ the limiting cases $\alpha_j \to 0$ and $\alpha_j \to \infty$ arise. The dependence parameters α_j, $j \in I$, are not easy to interpret. However, Coles and Tawn (1994) draw attention to the quantities $r_1 = (\alpha_i - \alpha_j)/2$ and $r_2 = (\alpha_i + \alpha_j)/2$ which can be interpreted as the asymmetry and intensity of the dependence between pairs of variables.

In this case, the exponent function can not be analytically computed, nonetheless it can still be evaluated numerically.

The second row of Figure 1.2 illustrates some examples of trivariate angular densities obtained with different sets of the parameters $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3)$. The plots from left to right have been obtained using the parameter sets $\{(2,2,2); (0.5,0.5,0.5); (2,2.5,30); (0.1,0.25,0.95)\}$. The first panel shows that, when values of the parameters are equal and greater than 1, the mass concentrates in the center of the simplex leading to strong dependence. The second panel shows the opposite, when values of $\boldsymbol{\alpha}$ are equal and less than 1, it yields to the case of weak dependence as the mass concentrates on the vertices of the simplex. The third panel shows the case of an asymmetric dependence structure and this is obtained when the values of the parameters are all greater than one. In this specific case the mass tends to spread towards the bottom and top left edges. The fourth panel illustrates another case of an asymmetric dependence structure, in this case obtained with all the values of the parameters that are less than 1, leading to a mass that concentrates along the top right edge and vertices.

1.3.3 Pairwise beta model

The tilted Dirichlet model has been successfully used for applications (e.g. Coles and Tawn, 1991), although it suffers from a lack of interpretability of the parameters. Cooley et al. (2010) proposed a similar model but with easily interpretable parameters. The definition of their model is based on a geometric approach. Specifically, they considered the symmetric pairwise beta function

$$h^*(w_i, w_j) = \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j}-1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j}-1}, i, j \in I,$$

where w_i and w_j are two elements of $\boldsymbol{w}$ and $\beta_{i,j} > 0$. This function has its center at the point $(1/d, \ldots, 1/d)$ and it verifies the first moment conditions (1.9). Then, the angular pairwise beta density is defined by summing together all the $d(d-1)/2$ possible pairs of variables, namely

$$h(\boldsymbol{w}; \boldsymbol{\theta}) = \frac{2(d-3)!\Gamma(\alpha d + 1)}{d(d-1)\Gamma(2\alpha+1)\Gamma\{\alpha(d-2)\}} \sum_{i,j \in I, i<j} h(w_i, w_j), \quad \boldsymbol{w} \in \mathbb{W},$$

where

$$h(w_i, w_j) = (w_i + w_j)^{2\alpha-1} \{1 - (w_i + w_j)\}^{\alpha(d-2)-d+2} h^*(w_i, w_j)$$

and $\boldsymbol{\theta} = (\alpha, \{\beta_{i,j}\}_{i,j \in I})$ with $\alpha > 0$. Each parameter $\beta_{i,j}$ controls the level of dependence between the i^{th} and the j^{th} components and the dependence increases for increasing values of $\beta_{i,j}$. The

function h^* is introduced to guarantee that the dependence ranges between weak and strong dependence. The parameter α controls the dependence of all the variables, when it increases the overall dependence increases.

Also in this case the exponent function can not be computed in closed form and hence it can only be evaluated numerically.

The third row of Figure 1.2 provides some examples of trivariate angular densities obtained with different values of the parameters $\boldsymbol{\theta} = (\alpha, \beta_{1,2}, \beta_{1,3}, \beta_{2,3})$. The plots from left to right have been obtained using the parameter sets $\{(4,2,2,2); (0.5,1,1,1); (1,2,4,15); (1,10,10,10)\}$. The first panel shows a case of symmetric density obtained with all equal parameters $\beta_{i,j}$ $i,j \in I$. A large value of the overall dependence parameter α pulls the mass towards the center of the simplex, indicating a strong dependence between the variables. On the contrary, the second panel shows that when the overall dependence parameter is close to zero then the mass concentrates on the vertices of the simplex, indicating weak dependence among the variables. The third panel illustrates a case of asymmetric angular density with strong dependence between the second and third variables that is due to a large value of $\beta_{2,3}$. Although the value of the global dependence parameter α is not large, it is enough to slightly push the mass towards the center of the simplex. The fourth panel shows a case of symmetric angular density, which is obtained with large values of the pairwise dependence parameters and an average value of the global dependence parameter. The mass is mainly concentrated on the center of the simplex and some mass tends to lie near the centers of the edges.

1.3.4 Hüsler-Reiss model

One of the most popular models is the Hüsler-Reiss (Hüsler and Reiss, 1989). Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n i.i.d. copies of a zero-mean unit variance Gaussian random vector. Assume that for all $i,j \in I$ the pairwise correlation $\rho_{i,j;n}$ satisfies the condition

$$\lim_{n\to\infty} \log n(1-\rho_{i,j;n}) = \lambda_{i,j}^2 \in [0,\infty).$$

Then, the exponent function of the limit distribution of $\boldsymbol{b}_n(\boldsymbol{M}_n - \boldsymbol{b}_n)$ for $n \to \infty$, where $\boldsymbol{b}_n = (b_n, \ldots, b_n)$ is a vector of real sequences (see Resnick, 2007, pp. 71-72), is

$$V(\boldsymbol{y};\boldsymbol{\theta}) = \sum_{j=1}^{d} \frac{1}{y_j} \Phi_{d-1}\Big\{\Big(\lambda_{i,j} + \frac{\log y_i/y_j}{2\lambda_{i,j}}\Big)_{i\in I_j}; \bar{\Lambda}_j\Big\}, \quad \boldsymbol{y} \in \mathbb{R}_+^d, \tag{1.20}$$

where $\boldsymbol{\theta} = \{\lambda_{i,j}\}_{i,j\in I}$, $I_j := I \setminus \{j\}$, Φ_{d-1} is $d-1$ dimensional Gaussian distribution with partial correlation $\bar{\Lambda}_j$. For all $j \in I$, the elements of $\bar{\Lambda}_j$ are $\lambda_{k,i;j} = (\lambda_{k,j}^2 + \lambda_{i,j}^2 - \lambda_{k,i}^2)/(2\lambda_{k,j}\lambda_{i,j})$, for $k,i \in I_j$. The parameter $\lambda_{i,j}$, $i,j \in I$, controls the dependence between the i^{th} and j^{th} elements of a vector of d extremes. These are completely dependent when $\lambda_{ij} = 0$ and become independent as $\lambda_{ij} \to \infty$.

In this case the angular measure concentrates on the interior of the simplex. Applying (1.11)

it can be checked (Engelke et al., 2012) that the angular density is

$$h(\boldsymbol{w};\boldsymbol{\theta}) = \phi_{d-1}\left\{\left(\lambda_{i,1} + \frac{\log w_i/w_1}{2\lambda_{i,1}}\right)_{i\in I_1}; \bar{\Lambda}_1\right\}\left\{w_1^2\prod_{i=2}^{d}(w_i 2\lambda_{i,1})\right\}^{-1}, \boldsymbol{w}\in\mathbb{W},$$

where ϕ_{d-1} is $d-1$ dimensional Gaussian density with partial correlation matrix $\bar{\Lambda}_1$.

The second last row of Figure 1.2 provides some examples of trivariate angular densities obtained with different values of the parameters $\boldsymbol{\theta} = (\lambda_{1,2}, \lambda_{1,3}, \lambda_{2,3})$. The plots from left to right have been obtained using the parameter sets $\{(0.3, 0.3, 0.3), (1.4, 1.4, 1.4), (1.7, 0.7, 1.1), (0.52, 0.71, 0.52)\}$. The first panel shows that with small and equal values of parameters the case of strong dependence among all the variables is obtained. In this case the mass concentrates around the center of the simplex. On the contrary, the second panel shows that with large and equal values of the parameters the case of weak dependence is obtained. In this case the mass is placed close to the vertices of the simplex. The third panel shows that an asymmetric dependence structure is obtained when the parameter values are different. In this case the mass tends to concentrate around the vertices and edges that are concerned with the smaller values of the parameters. The fourth panel shows that a symmetric dependence structure, with respect to the second component is obtained setting the values of two parameters to be equal. In this case the mass is equally divided up towards the two vertices and edges that are concerned with the smaller values of the parameters.

1.3.5 Extremal-t model

The extremal-t model (Nikoloulopoulos et al., 2009) is more flexible than the Hüsler-Reiss but it is still simple enough. It is easily interpretable and useful in practical applications (see Davison et al., 2012). Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n i.i.d. copies of a zero-center unit scale Student-t random vector with dispersion matrix Σ and $\nu > 0$ degrees of freedom (d.f.). Then, the exponent function of the limiting distribution of $\boldsymbol{M}_n/\boldsymbol{a}_n$ for $n\to\infty$, where $\boldsymbol{a}_n = (a_n \ldots, a_n)$ is a vector of positive sequences (see Demarta and McNeil, 2005), is

$$V(\boldsymbol{y};\boldsymbol{\theta}) = \sum_{j=1}^{d}\frac{1}{y_j}T_{d-1,\nu+1}\left\{\left[\sqrt{\frac{\nu+1}{1-\rho_{i,j}^2}}\left\{\left(\frac{y_i}{y_j}\right)^{\frac{1}{\nu}} - \rho_{i,j}\right\}\right]_{i\in I_j}; \bar{\Sigma}_j\right\}, \tag{1.21}$$

for all $\boldsymbol{y}\in\mathbb{R}_+^d$, where $\boldsymbol{\theta} = (\{\rho_{i,j}\}_{i,j\in I}, \nu)$ and $T_{d-1,\nu+1}$ is a $d-1$ dimensional Student-t distribution with $\nu+1$ d.f. and partial correlation matrix $\bar{\Sigma}_j$. The correlation parameter $\rho_{i,j}$, $i,j\in I$, drives the dependence between pairs of variables with the dependence that increases with the increasing of $\rho_{i,j}$. The parameter ν controls the overall dependence among all the variables. For decreasing values of ν the dependence increases and vice versa.

The Hüsler-Reiss model is a special case of the extremal-t. Indeed, for all $i,j\in I$ if the correlation parameters of the extremal-t distribution are equal to $\rho_{i,j;\nu} = 1-\lambda_{i,j}^2/\nu$, then this distribution converges weakly, as $\nu\to\infty$, to the Hüsler-Reiss (see Nikoloulopoulos et al., 2009).

In this case the angular measure places mass on all the subspaces of the simplex. When

$\mathcal{S} = I$, then applying (1.11) we obtain that the angular density is

$$h(\boldsymbol{w};\boldsymbol{\theta}) = \frac{t_{d-1,\nu+1}\left(\left[\sqrt{\frac{\nu+1}{1-\rho_{i,1}^2}}\left\{(w_i/w_1)^{1/\nu} - \rho_{i,1}\right\}\right]_{i\in I_1}; \bar{\Sigma}_1\right)}{\nu^{d-1} w_1^{d+1}\left\{\prod_{i=2}^{d}\sqrt{\frac{\nu+1}{1-\rho_{i,1}^2}}\,(w_i/w_1)^{(\nu-1)/\nu}\right\}^{-1}}, \quad \boldsymbol{w}\in\mathbb{W},$$

where $t_{d-1,\nu+1}$ is $d-1$ dimensional Student-t density with partial correlation matrix $\bar{\Sigma}_1$ (e.g. Ribatet, 2013). When $\mathcal{S} = \{j\}$, then applying (1.12) we obtain that the mass on the extreme points of the simplex is

$$h_{d,\mathcal{S}} = T_{d-1,\nu+1}\left[\left\{-\rho_{i,j}(\nu+1)^{1/2}/(1-\rho_{i,j}^2)^{1/2}\right\}_{i\in I_j}; \bar{\Sigma}_j\right], \quad j\in I.$$

The last row of Figure 1.2 provides some examples of the trivariate angular densities obtained with different values of the parameters $\boldsymbol{\theta} = (\rho_{1,2}, \rho_{1,3}, \rho_{2,3}, \nu)$. From left to right the plots are obtained using the parameter values $\{(0.95, 0.95, 0.95, 2); (-0.3, -0.3, -0.3, 5); (0.52, 0.71, 0.52, 3); (0.52, 0.71, 0.52, 2)\}$. The first panel shows that when the scale parameters ρ_{ij} are all equal and close to one and the d.f. ν are small, then the mass concentrates around the center of the simplex and therefore the dependence is strong. The second panel shows the opposite, when the correlations are close to zero and the d.f. are high, the mass concentrates around the vertices of the simplex and hence the dependence is weak. The third panel shows that when two scale parameters are equal then the dependence structure is symmetric with respect to the second component and the mass tends to concentrate on the top vertex and the bottom edge and vertices. The fourth panel shows that with the same setting but with smaller d.f. the mass is pushed towards the center of the simplex and hence the dependence is stronger.

1.4 Estimating the extremal dependence

Several inferential methods have been explored for inferring the extremal dependence. Nonparametric and parametric approaches are available. In the first case recent advances are Gudendorf and Segers (2011), Gudendorf and Segers (2012) and Marcon et al. (2014), see also the references therein. Both likelihood based and Bayesian inferential methods have been widely investigated. Examples of likelihood based methods are the approximate likelihood (e.g. Coles and Tawn, 1994; Cooley et al., 2010; Engelke et al., 2012) and the composite likelihood (e.g. Padoan et al., 2010; Davison and Gholamrezaee, 2012). Examples of Bayesian techniques are Apputhurai and Stephenson (2011), Sabourin et al. (2013), Sabourin and Naveau (2014).

For comparison purposes in the next section the real data analysis is performed using the maximum approximate likelihood estimation method and the approximate Bayesian method based on the approximate likelihood. Here is a brief description.

From the theory in Sections 1.2.1, if $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are i.i.d. copies of $\boldsymbol{Y}$ on $\mathbb{R}_+^d$ with a distribution in the domain of attraction of a MEVD, then the distribution of the sequence

$\{R_i/n, \boldsymbol{W}_i, i = 1, \dots, n\}$, where $R_i = Y_{i,1} + \cdots + Y_{i,d}$ and $\boldsymbol{W}_i = \boldsymbol{Y}_i/R_i$, converges as $n \to \infty$ to the distribution of a PPP with density $\mathrm{d}\psi(r, \boldsymbol{w}) = r^{-2}\mathrm{d}r \times \mathrm{d}H(\boldsymbol{w})$.

Assume that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are i.i.d. observations from a random vector with an unknown distribution. Since the aim is estimating the extremal dependence, we transform the data into the sample $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ with unit Fréchet marginal distributions. This is done by applying the probability integral transform, after fitting the marginal distributions. Next, the coordinates of the data-points are changed from Euclidean into pseudo-polar by the transformation

$$r_i = y_{i,1} + \cdots + y_{i,d} \quad \boldsymbol{w}_i = \boldsymbol{y}_i/r_i, \quad i = 1, \dots, n.$$

Then, the sequence $\{(r_i, \boldsymbol{w}_i), i = 1, \dots, n : r_i > r_0\}$, where $r_0 > 0$ is a large threshold, comes approximately from a Poisson point process with intensity measure ψ. Let $\mathcal{W}_{r_0} = \{(r, \boldsymbol{w}) : r > r_0\}$ be the set of points with a radial component larger than r_0, then the number of points falling in $\mathcal{W}_{r_0}$ is given by $N(\mathcal{W}_{r_0}) \sim \mathrm{Pois}\{1/\psi(\mathcal{W}_{r_0})\}$. Conditionally to $N(\mathcal{W}_{r_0}) = m$, the points $\{(r_{(i)}, \boldsymbol{w}_{(i)}), i = 1, \dots, m\}$ are i.i.d. with common density $\mathrm{d}\psi(r, \boldsymbol{w})/\psi(\mathcal{W}_{r_0})$. If we assume that H is known apart from a vector of unknown parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, then the approximate likelihood of the excess is

$$\begin{aligned} L(\boldsymbol{\theta}; (r_{(i)}, \boldsymbol{w}_{(i)}), i = 1, \dots, m) &= \frac{e^{-\psi(\mathcal{W}_{r_0})}\psi(\mathcal{W}_{r_0})^m}{m!} \prod_{i=1}^{m} \frac{\mathrm{d}\psi(r_{(i)}, \boldsymbol{w}_{(i)})}{\psi(\mathcal{W}_{r_0})} \\ &\propto \prod_{i=1}^{m} h(\boldsymbol{w}_{(i)}, \boldsymbol{\theta}), \end{aligned} \tag{1.22}$$

where h is a parametric angular density function (e.g. Engelke et al., 2012; Beirlant et al., 2006, pp. 170–171). In the next section the angular density models described in Section 1.3 are fitted to the data by the maximization of the likelihood (1.22). For brevity the asymmetric logistic model is not considered since it has too many parameters. The likelihood (1.22) is proportional to the product of angular densities, therefore the maximizer of (1.22) is obtained equivalently by maximizing the log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \log h(\boldsymbol{w}_{(i)}, \boldsymbol{\theta}). \tag{1.23}$$

Denote by $\widehat{\boldsymbol{\theta}}$ the maximizer of ℓ and by $\ell'(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta})$ the score function. Since (1.22) provides an approximation of the true likelihood, then from the theory on model misspecification (e.g. Davison, 2003, pp. 147–148) it follows that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_p(\mathbf{0}, J(\boldsymbol{\theta})^{-1}\, K(\boldsymbol{\theta})\, J(\boldsymbol{\theta})^{-1}), \quad n \to \infty,$$

where $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ is the p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ, $\boldsymbol{\theta}$ is the true parameter and

$$J(\boldsymbol{\theta}) = -\,\mathbb{E}\{\nabla_{\boldsymbol{\theta}} \ell'(\boldsymbol{\theta})\}, \qquad K(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}\{\ell'(\boldsymbol{\theta})\},$$

are the sensitive and variability matrices (Varin et al., 2011). In the case of misspecified models, model selection can be performed by computing the Takeuchi Information Criterion (TIC) (e.g. Sakamoto et al., 1986), that is

$$TIC = -2\left[\ell(\widehat{\boldsymbol{\theta}}) - \mathrm{tr}\{K(\widehat{\boldsymbol{\theta}})J^{-1}(\widehat{\boldsymbol{\theta}})\}\right],$$

where the log-likelihood, the variability and sensitive matrices are evaluated at $\widehat{\boldsymbol{\theta}}$. The model with the smallest value of the TIC is preferred.

In order to derive an approximate posterior distribution for the parameters of an angular density, the approximate likelihood (1.22) can be used within the Bayesian paradigm (see Sabourin et al., 2013). Briefly, let $q(\boldsymbol{\theta})$ be a prior distribution on $\boldsymbol{\theta}$, then the posterior distribution of the angular density's parameters is

$$q(\boldsymbol{\theta}|\boldsymbol{w}) = \frac{\prod_{i=1}^{m} h(\boldsymbol{w}_{(i)}, \boldsymbol{\theta})\, q(\boldsymbol{\theta})}{\int_{\Theta} \prod_{i=1}^{m} h(\boldsymbol{w}_{(i)}, \boldsymbol{\theta})\, q(\boldsymbol{\theta})\, \mathrm{d}\,\boldsymbol{\theta}}. \tag{1.24}$$

With the angular density models in Section (1.3) the analytical expression of $q(\boldsymbol{\theta}|\boldsymbol{w})$ can not be derived. Therefore, we use a Markov Chain Monte Carlo method for sampling from an approximation of $q(\boldsymbol{\theta}|\boldsymbol{w})$. Specifically, we use a Metropolis–Hastings simulating algorithm (e.g. Hastings, 1970). With the pairwise beta models we use the prior distributions described by Sabourin et al. (2013). With the tilted Dirichlet and Hüsler-Reiss model we use independent zero-mean normal prior distributions with standard deviations equal to 3 for $\log \alpha_j$ and $\log \lambda_{i,j}$ with $i, j \in I$. For the extremal-t model we use independent zero-mean normal prior distributions with standard deviations equal to 3 for $\mathrm{sign}(\rho_{ij})\mathrm{logit}(\rho_{ij}^2)$ with $i, j \in I$, where $\mathrm{sign}(x)$ is the sign of x for $x \in \mathbb{R}$ and $\mathrm{logit}(x) = \log(x/(1-x))$ for $0 \leq x \leq 1$, and a zero-mean normal prior distribution with standard deviations equal to 3 for $\log \nu$. Similar to Sabourin et al. (2013), for each models' parameter we select a sample of 50×10^3 observations from the approximate posterior, after a burn-in period of length 30×10^3. These sizes have been determined using the Geweke convergence diagnostics (Geweke, 1992) and the Heidelberger and Welch test (Heidelberger and Welch, 1981) respectively.

Model selection is performed using the Bayesian Information Criterion (BIC) (e.g. Sakamoto et al., 1986), that is

$$BIC = -2\,\ell(\widehat{\boldsymbol{\theta}}) + p\{\log m + \log(2\pi)\},$$

where p is the number of parameters and m is the sample size. The model with the smallest value of the BIC is preferred.

1.5 Real data analysis: Air quality data

We analyze the extremal dependence of the air quality data, recorded in the city centre of Leeds, UK. The aim is to estimate the probability that multiple pollutants will be simultaneously

high in the near future. This dataset has been previously studied by Heffernan and Tawn (2004), Boldi and Davison (2007) and Cooley et al. (2010). The data are the daily maximum of five air pollutants: particulate matter (PM10), nitrogen oxide (NO), nitrogen dioxide (NO2), ozone (03), and sulfur dioxide (SO2). Levels of the gases are measured in parts per billion, and those of PM10 in micrograms per cubic meter. We focus our analysis on the winter season (from November to February) from 1994 to 1998.

A preliminary analysis focuses on the data of triplets of variables. For brevity we only report the results of the most dependent triplets: PM10, NO, SO2 (PNS), NO2, SO2, NO (NSN) and PM10, NO, NO2 (PNN). For each variable, the empirical distribution function is estimated with the data below the 0.7 quantile and a GPD is fitted to the data above the quantile (Cooley et al., 2010). Then, each marginal distribution is transformed into a unit Fréchet. The coordinates of the data-points are transformed to radial distances and angular components. For each triplet, the 100 observations with the largest radial distances are retained. The angular density models in Section 1.3 are fitted to the data using the methods in Section 1.4.

The results are presented in Table 1.1. Maximum likelihood estimates are similar to the estimated posterior means and the estimated posterior standard deviations are typically larger than the standard errors. For PNS we obtain the same maximum likelihood estimates as Cooley et al. (2010) with the pairwise beta model, however we use (1.4) to compute the variances of the estimates and so we attain larger standard errors than they do. Both the TIC and BIC lead to the same model selection. The Hüsler-Reiss model provides the best fit for all the groups of pollutants.

From top to bottom, Figure 1.3 displays the angular densities, computed with the posterior means. From left to right the Hüsler-Reiss, the tilted Dirichlet and the pairwise beta densities are reported. With PNS, we see that there are many observations along the edge that link PM10 and NO, revealing strong dependence between these two pollutants. There are also several observations on the SO2 vertex, reflecting that this pollutant is mildly dependent with the other two. There are also some data in the middle of the simplex, indicating that there is mild dependence among the pollutants. Similarly, with NSN we see that there is strong dependence between NO and NO2, because there are many observations along the edge that link them. There is a mild dependence between SO2 and the other pollutants, because there is a considerable amount of data on the O3 vertex. Overall, there is mild dependence among the pollutants, because there is a small amount of data in the middle of the simplex. With PNN we see that most of the observations are placed on the middle of the simplex revealing an overall strong dependence among the pollutants. There is a small amount of data along the edge that link NO2 and NO and on the PM10 vertex. This reflects more dependence between NO2 and NO than between NO2 and PM10 and PM10 and NO. All these features are well captured by the angular densities estimated using the Hüsler-Reiss model.

With this analysis we found that O3 is only weakly dependent with the other pollutants. This result was also found by Heffernan and Tawn (2004). Then, the second part of the analysis focuses only on PM10, NO, NO2 and SO2. Now, because a larger number of parameters needs to be estimated, then the 200 observations with the largest radial distances are selected (see

Model	Method	Estimates				$\ell(\widehat{\boldsymbol{\theta}})$	TIC/BIC
TD		$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$			
PNS	L	1.20(0.24)	0.67(0.07)	0.41(0.08)		199.63	−399.21
	B	1.22(0.25)	0.68(0.11)	0.42(0.09)			−379.90
NSN	L	0.85(0.12)	0.39(0.08)	0.90(0.11)		200.84	−401.63
	B	0.86(0.15)	0.39(0.09)	0.81(0.15)			−382.32
PNN	L	1.43(0.28)	1.55(0.31)	1.28(0.20)		186.35	−372.64
	B	1.45(0.30)	1.57(0.28)	1.29(0.23)			−353.36
PB		$\widehat{\beta}_{1,2}$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{2,3}$	$\widehat{\alpha}$		
PNS	L	3.21(0.70)	0.47(0.05)	0.45(0.04)	0.68(0.06)	95.95	−191.87
	B	3.31(1.13)	0.48(0.11)	0.46(0.10)	0.68(0.09)		−166.10
NSN	L	0.40(0.03)	3.74(1.77)	0.50(0.05)	0.64(0.05)	102.59	−205.13
	B	0.40(0.09)	4.00(1.72)	0.51(0.12)	0.64(0.08)		−179.36
PNN	L	3.75(1.38)	0.71(0.09)	3.18(1.21)	1.35(0.18)	84.31	−168.55
	B	3.83(1.75)	0.72(0.16)	3.70(1.80)	1.37(0.20)		−142.66
HR		$\widehat{\lambda}_{1,2}$	$\widehat{\lambda}_{1,3}$	$\widehat{\lambda}_{2,3}$			
PNS	L	0.65(0.06)	0.90(0.04)	0.98(0.03)		234.51	−468.93
	B	0.65(0.04)	0.90(0.04)	0.98(0.04)			−449.67
NSN	L	1.00(0.04)	0.56(0.04)	0.96(0.04)		251.80	−503.54
	B	1.00(0.04)	0.57(0.03)	0.97(0.04)			−484.25
PNN	L	0.60(0.05)	0.70(0.04)	0.51(0.03)		198.23	−396.38
	B	0.60(0.03)	0.70(0.04)	0.51(0.03)			−377.11
ET		$\widehat{\rho}_{1,2}$	$\widehat{\rho}_{1,3}$	$\widehat{\rho}_{2,3}$	$\widehat{\nu}$		
PNS	L	0.87(0.02)	0.74(0.03)	0.66(0.03)	3.89(0.51)	152.13	−304.18
	B	0.87(0.02)	0.77(0.02)	0.72(0.01)	4.02(0.35)		−275.13
NSN	L	0.58(0.04)	0.87(0.02)	0.64(0.03)	3.50(0.01)	141.92	−283.80
	B	0.72(0.01)	0.89(0.02)	0.73(0.02)	4.00(0.33)		−242.50
PNN	L	0.88(0.02)	0.82(0.02)	0.89(0.01)	3.70(0.78)	180.74	−361.38
	B	0.86(0.02)	0.78(0.03)	0.87(0.02)	3.21(0.43)		−330.33

Table 1.1: Summary of the extremal dependence models fitted to the UK air pollution data. For each angular density model the estimation results of the triplets of pollutants are reported. L and B denote the approximate likelihood and Bayesian inferential method. Estimates are maximum likelihood (standard errors) and posterior means (standard deviations).

	Tilted Dirichlet	Pairwise Beta	Hüsler-Reiss	extremal-t
$\ell(\widehat{\boldsymbol{\theta}})$	654.3	402.5	762.7	532.3
TIC	−1308.6	−805.0	−1525.3	−1064.5
BIC	−1280.0	−753.4	−1475.5	−974.7

Table 1.2: Summary of the extremal dependence models fitted to the UK air pollution data.

Cooley et al., 2010). Table 1.2 presents the estimation results. For brevity we only report the maximum value of the log-likelihood, the TIC and the BIC. The Hüsler-Reiss model provides the smallest values of the TIC and BIC, revealing again that it better fits the pollution data. Accordingly hereafter calculations will be made using this model and the estimates obtained

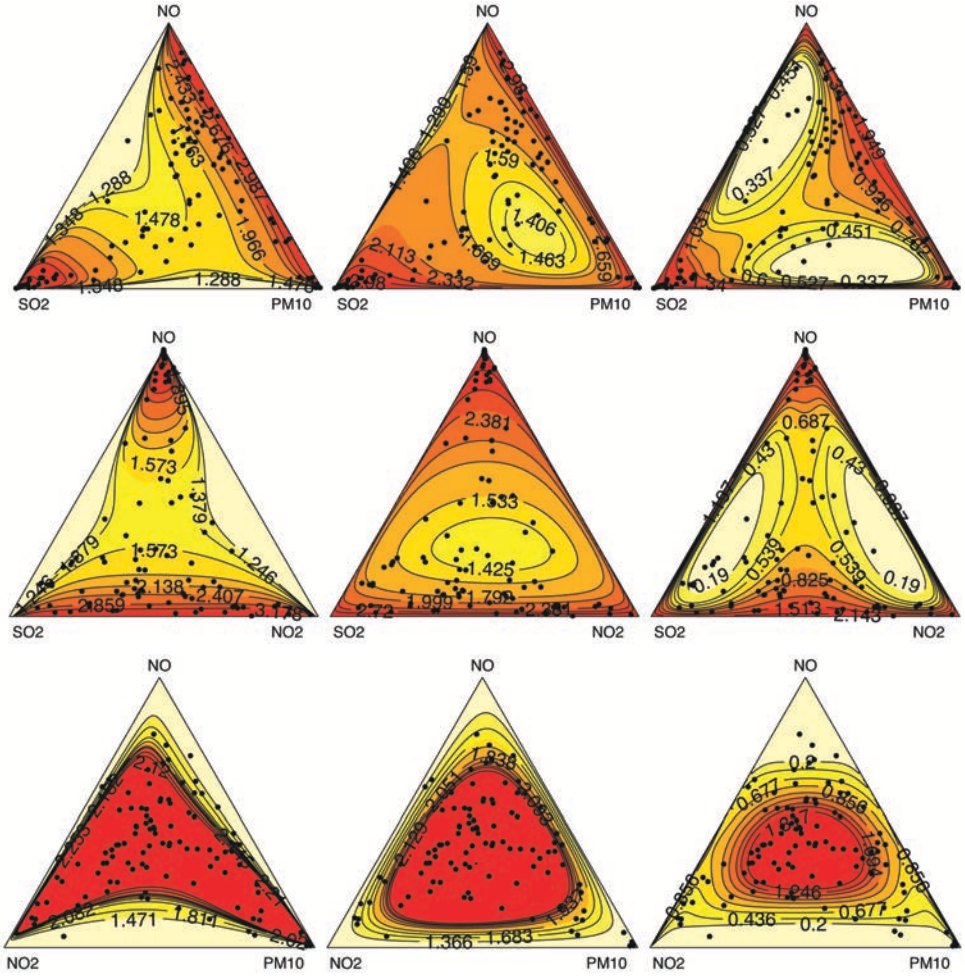

Figure 1.3: Estimated angular densities in logarithm scale. Dots represent the largest 100 observations.

with the Bayesian approach.

We summarize the extremal dependence of the four variables using the extremal coefficient (1.13) and the coefficient of tail dependence (1.16). Specifically, $\widehat{\vartheta} = 2.267$ with a 95% credible interval is equal to $(1.942, 2.602)$ and $\widehat{\chi} = 0.242$ with a 95% credible interval is $(0.150, 0.361)$. These results suggest a strong extremal dependence among the pollutants. The estimated extremal dependence can be used in turn to estimate the probability that multiple pollutants exceed a high threshold. Consider a value $\boldsymbol{y}$ whose radial component is a high threshold r_0. Then, the probability of falling in the failure region (1.14) is approximately equal to the right hand side of (1.18). Because the exponent function is related to the tail function by the inclusion-exclusion principle, then using (1.20) we have

$$\text{pr}\{Y_1 > y_1, \ldots, Y_d > y_d\} \approx \sum_{j=1}^{d} \frac{1}{y_j} \bar{\Phi}_{d-1}\Big\{\Big(\lambda_{k,j} + \frac{\log y_k/y_j}{2\lambda_{k,j}}\Big)_{k \in I_j}; \bar{\Lambda}_j\Big\}, \tag{1.25}$$

	Event 1	Event 2	Event 3
Excess / n	18/528	14/562	12/528
Emp. Est.	0.034 (0.019, 0.050)	0.025 (0.012, 0.038)	0.023 (0.010, 0.035)
Mod. Est.	0.038	0.030	0.030

Table 1.3: Probability estimates of excesses. The first row reports the number of excess and the sample size. The second row reports the empirical estimates and between brackets the 95% confidence intervals obtained with the normal approximation. The third row reports the model estimates.

where $\bar{\Phi}_{d-1}$ is the survival function of the multivariate normal distribution (Nikoloulopoulos et al., 2009). Similar to Cooley et al. (2010) we define three extreme events: $\{\text{PM10} > 95, \text{NO} > 270, \text{SO2} > 95\}$, $\{\text{NO2} > 110, \text{SO2} > 95, \text{NO} > 270\}$ and $\{\text{PM10} > 95, \text{NO} > 270, \text{NO2} > 110, \text{SO2} > 95\}$. Then, we compute probability (1.25) using in place of the parameters their estimates. Table 1.3 reports the results. For the three events the estimates fall inside the 95% confidence intervals highlighting the ability of the model to estimate such extreme events.

The right-hand side of (1.25) can also be used for estimating joint return levels. In the univariate case see Coles (2001, pp.49–50). In the multivariate case different definitions of return levels may be available (Johansen, 2004). Let $J \subset I$, $\{x_i, i \in I \backslash J\}$ be a sequence of fixed high thresholds and $p \in (0, 1)$ be a fixed probability. Given a return period $1/p$, we define *joint*

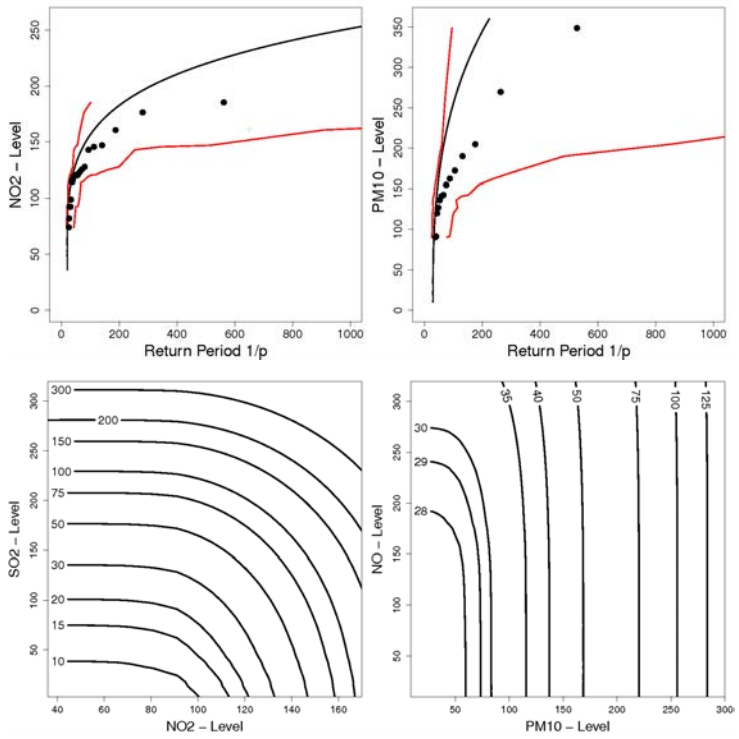

Figure 1.4: Joint return level plots of single components NO2 and PM10 and of the two components (SO2, NO2) and (NO, PM10)

return levels the quantiles $\{y_{j;p}, j \in J\}$ that satisfy the equation

$$p = \text{pr}(Y_j > y_{j;p}, Y_i > x_i,\ j \in J, i \in I \backslash J).$$

Figure 1.4 displays univariate and bivariate *joint return level plots*. When $J = \{j\}$, with $j \in I$, then the joint return level plot displays $y_{j;p}$ against $1/p$ for different values of p. When $J = \{i, j\}$, with $i, j \in I$, then for different values of $1/p$ the contour levels of $(y_{i;p}, y_{j;p})$ are displayed. With solid lines, the top-left and right panels of Figure 1.4 report the estimated return levels of NO2 and PM10 jointly to the extreme events $\{\text{SO2} > 95, \text{NO} > 270\}$ and $\{\text{NO} > 270, \text{NO2} > 110, \text{SO2} > 95\}$ respectively. The dots are the empirical estimates and the red solid lines are the pointwise 95% confidence intervals. These are computed using the normal approximation when $p > 0.02$ and using exact binomial confidence intervals when $p < 0.02$. The bottom-left and right panels report the contour levels of the return levels for (NO2, SO2) and (PM10, NO) jointly to events $\{\text{NO} > 270\}$ and $\{\text{NO2} > 110, \text{SO2} > 95\}$ respectively.

The joint return level can be interpreted as follows. For example, from the top-right panel we have that the 50 years joint return level of PM10 is 166. Concluding, we expect that PM10 will exceed the level 166 together with the event that NO, SO2 and NO2 simultaneously exceed the levels 270, 95 and 110 respectively, on average every 50 years.

1.6 Computational details

The figures and the estimation results have been obtained using the free software `R` (Team, 2013) and in particular the package `ExtremalDep`, available at `https://r-forge.r-project.org/projects/extr`
Bayesian estimation is obtained using and extending some routines of the package `BMAmev`. The left and middle panels of Figure 1.1 were obtain using the routines `scatter3d` and `polygon3d` of the package `plot3D`.

Chapter 2

Max-stable Processes

2.1 Introduction

In the previous chapter, multivariate extreme value theory was applied to pollution data from the city of Leeds, UK, in order to analyze the extremal dependence between five different air pollutants. The goal was to evaluate the probability of some air pollutant levels to be concurrently large, which may have a strong negative impact on the population and its environment. However some environmental phenomena are spatial by nature, which brings additional challenges that cannot be resolved by the multivariate analysis of extremes only. This justifies the use of stochastic processes and the analysis of their extreme behaviour as it generalizes multivariate analysis of extremes to a whole area of interest. This topic is considered as the study of infinite-dimensional extremes. For example, consider the temperature over a region $\mathbb{S}$ defined as $\{X(s)\}_{s\in\mathbb{S}}$ and suppose $X_1, X_2, \ldots$ i.i.d. replicates of X representing the daily maxima (minima) temperature across $\mathbb{S}$. It might be of interest to predict where, on $\mathbb{S}$, extreme temperatures are expected to happen and what maximum (minimum) value they could reach. For example, this can be summarized by the study the extremal behaviour of the aggregate process

$$\int_{\mathbb{S}} X(s)\mathrm{d}s,$$

or, under the assumption that there is a large number n of observations of the process, by the estimation of a probability to exceed a deterministic function t across space such as

$$\Pr(X(s) > t(s), \forall s \in \mathbb{S}).$$

The function t represents the critical temperature level over $\mathbb{S}$. Typically, authorities often define a heat wave when the minimum temperature at night exceeds some high threshold (given by t) for more than three consecutive days. It is most likely that the observed processes don't reach the critical level of temperature t causing trouble when estimating the tail distribution of the processes. Limit theory is then introduced in order to bypass this problem, i.e. we consider the case where $n \to \infty$. The assumption that the observed temperatures are far from the critical level, is still required in the limiting case. This means that t is actually a function of n that

approaches the upper bound of the distribution of X as n gets large and it leads us to study the limit theory for the pointwise maxima of i.i.d. random processes.

This limit theory will be introduced on the set $[0,1]$ for convenience but most results can be applied to any compact Euclidean space. We define by $\mathrm{C}[0,1]$ the space of all continuous functions on $[0,1]$ equipped with the uniform metric $d_\infty(f,g) \equiv \|f-g\|_\infty = \sup_x |f(x)-g(x)|$. Since the set of continuous functions on $[0,1]$ is bounded, then $(\mathrm{C}[0,1], d_\infty)$ is a metric space that is separable and complete.

Throughout this chapter, we consider a stochastic process X on $\mathrm{C}[0,1]$, meaning that it has continuous sample paths, defined by $\{X(s)\}_{s\in[0,1]}$ with continuous marginal distribution function $F_s(x) = \Pr(X(s) \leq x)$. Assume also that there exists normalizing constants given as two continuous functions of s, $a_n(s)$ positive and $b_n(s)$ such that

$$\{\eta_n(s)\}_{s\in[0,1]} := \left\{ \max_{1\leq i\leq n} \frac{X_i(s) - b_n(s)}{a_n(s)} \right\}_{s\in[0,1]} \xrightarrow{\mathcal{D}} \{\tilde{\eta}(s)\}_{s\in[0,1]}, \tag{2.1}$$

in $C[0,1]$, where $X_1, X_2, \ldots$ are i.i.d. replicates of X and $\tilde{\eta}(s)$ has non-degenerate marginal distributions. We denote the marginal distribution by $G_s(x) = \Pr(\tilde{\eta}(s) \leq x)$. Note that if the process X behaves badly, there might not exists normalizing constants $a_n(s)$ and $b_n(s)$ such that the rescaled maxima converges weakly in $\mathrm{C}[0,1]$.

The notation $\xrightarrow{\mathcal{D}}$ defines the convergence in distribution or weak convergence in a space to be specified. In (2.1), convergence occurs in the $\mathrm{C}[0,1]$ space, as the maximum of continuous stochastic processes, rescaled by continuous functions, is still a continuous process. Note the analogy with the multivariate case given in (1.1). Taking the temperature field example mentioned above with $\mathcal{S} = [0,1]$, the replicates $X_i(s)$ of the process $\{X(s)\}_{s\in[0,1]}$ can be seen as the temperatures observed at time i at a location s in $[0,1]$.

Remark 2.1. Max-stable processes are built on the fundamental assumption (2.1) which relies on the notion of weak convergence in $\mathrm{C}[0,1]$. Billingsley (1968, Theorem 8.1) states that the process η_n converges in $\mathrm{C}[0,1]$ if its finite-dimensional distribution converges and the tightness criterion is met. Tightness is proved using Prohorov's theorem applied to the characterization of compact sets given by Arzelá-Ascoli theorem.

As explained in Chapter 1, the main goal is to derive the dependence structure and a common choice is to standardize the margins in order to be only focusing on this structure. This standardization procedure of the marginal distributions is detailed in Section 2.2. Some theoretical results on the limiting distribution are presented, leading to the definition of max-stable processes. The specification of the margins helps to define, in Section 2.3, the exponent measure which is used to characterize the dependence structure. Further, two characterizations of these max-stable processes will be given as well as numerical summaries to assess the quantity of spatial dependence (Section 2.4 and 2.5). Finally some parametric models will be listed in Section 2.6 and their characteristics highlighted.

2.2 On the form of the marginal distributions

In section 2.1 we assumed that the marginal distributions of X were continuous and given by F_s. Also, the replicates $X_1, X_2, \ldots$ have the same marginal distributions. If there are sequences of continuous functions such that the limiting result (2.1) occurs, then such convergence leads to convergence also of the marginal distributions, that is

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(\max_{1\leq i\leq n} \frac{X_i(s) - b_n(s)}{a_n(s)} \leq x\right) &= \lim_{n\to\infty} \Pr\left(\frac{X_1(s) - b_n(s)}{a_n(s)} \leq x, \ldots, \frac{X_n(s) - b_n(s)}{a_n(s)} \leq x\right) \\ &= \lim_{n\to\infty} \Pr\left(\frac{X(s) - b_n(s)}{a_n(s)} \leq x\right)^n \\ &= \lim_{n\to\infty} F_s^n\left(a_n(s)x + b_n(s)\right) \\ &= G_s(x) \equiv \Pr\left(\tilde{\eta}(s) \leq x\right) \end{aligned} \tag{2.2}$$

uniformly for $s \in [0,1]$. If the convergence results (2.1) and (2.2) hold, then for all fixed $s \in [0,1]$, G_s must be a GEV distribution by construction. It is given by

$$G_s(x) = \exp\left\{-\left(1 + \xi(s)\frac{(x-\mu(s))}{\sigma(s)}\right)^{-1/\xi(s)}\right\}, \text{ with } 1 + \xi(s)\frac{(x-\mu(s))}{\sigma(s)} > 0. \tag{2.3}$$

When the limiting result applies, we say that F is in the maximum domain of attraction of G.

Define two extreme value cdfs G and $G^\star$ as being of the same type if $G^\star(x) = G(ax+b)$, for some $a > 0$ and b and for all x and where G is a Generalized Extreme Value (GEV) distribution (Fisher and Tippett, 1928; Gnedenko, 1943). Using the parametrization of the extremal types given by Misés (1936) and Jenkinson (1955) we consider that the Type I, II and III distributions respectively refer to the Gumbel, Fréchet and Weibull distributions.

Without loss of generality, we assume for simplicity that the location $\mu(s)$ and scale $\sigma(s)$ functions are equal to the constants 0 and 1 for all $s \in [0,1]$. Consequently, taking the logarithm of (2.2) with the GEV assumption and using a first order Taylor expansion we obtain

$$\lim_{n\to\infty} n\left\{1 - F_s(a_n(s)x + b_n(s))\right\} = (1 + \xi(s)x)^{1/\xi(s)}, \tag{2.4}$$

where ξ is a continuous function of s called the index function. Due to the monotonicity property of distribution functions, both left and right hand side of (2.4) are monotonic functions and hence have an inverse. Let U_s be the left-continuous inverse of $1/(1-F_s)$, we can provide a similar uniform convergence result as (2.4) for $s \in [0,1]$ considering the inverse functions. Then for $u \in (0,\infty)$, this yields of

$$\lim_{n\to\infty} \frac{U_s(nu) - b_n(s)}{a_n(s)} = \frac{u^{\xi(s)} - 1}{\xi(s)}. \tag{2.5}$$

Thus, an equivalence arises between (2.1), (2.5) and

$$\left\{\max_{1\leq i\leq n}\frac{1}{n\{1-F_s(X_i(s))\}}\right\}_{s\in[0,1]}\xrightarrow{\mathcal{D}}\left\{(1+\xi(s)\tilde{\eta}(s))^{1/\xi(s)}\right\}_{s\in[0,1]}.$$

Set $\{\eta\}_{s\in[0,1]}:=\{(1+\xi(s)\tilde{\eta}(s))^{1/\xi(s)}\}_{s\in[0,1]}$, and denote by $G_{s,0}(x)=Pr(\eta(s)\leq x)$, then for all fixed s in $[0,1]$, $G_{s,0}(x)$ is a unit Fréchet distribution. Furthermore, a stochastic process Z on $\mathrm{C}[0,1]$, with non-degenerate margins, is said to be max-stable if there exist continuous normalizing functions $a_n(s)>0$ and $b_n(s)$ such that

$$\{Z(s)\}_{s\in[0,1]}\overset{\mathcal{D}}{=}\left\{\max_{1\leq i\leq n}\frac{Z_i(s)-b_n(s)}{a_n(s)}\right\}_{s\in[0,1]},$$

where Z_i are independent replicates of Z. de Haan (1984) proved that when the limiting process in (2.1) exists then this must be a max-stable process.

Hereafter we assume to work with the process $\eta(s)$ that has unit Fréchet marginal distributions which is often named simple max-stable process. Hence the normalizing constants are independent of s and, applying a classical result from univariate extreme value theory, given by $a_n(s)=n$ and $b_n(s)=0$.

2.3 The dependence structure

According to relation (2.5) we can focus only on positive functions in $\mathrm{C}[0,1]$. Hence, in this section, we investigate the dependence structure of simple max-stable processes based on this restriction. The link with the analysis of multivariate extremes dependence from Chapter 1 will be highlighted.

Denote by $\mathrm{C}^+[0,1]$ the set of strictly positive functions that belongs to $\mathrm{C}[0,1]$. Set the normalizing functions $a_n(s)=n$ and $b_n(s)=0$, and let $Y_1,Y_2,\ldots$ be i.i.d. copies of a stochastic processes Y in $\mathrm{C}^+[0,1]$ and in the domain of attraction of a simple max-stable process, that is

$$\left\{\frac{1}{n}\max_{1\leq i\leq n}Y_i(s)\right\}_{s\in[0,1]}\xrightarrow{\mathcal{D}}\{\eta(s)\}_{s\in[0,1]},\tag{2.6}$$

in $\mathrm{C}^+[0,1]$, where η is unit Fréchet distributed for $s\in[0,1]$ and thus is simple max-stable.

When (2.6) occurs then the following facts follow. Suppose $\mathcal{B}$ a Borel set of $\mathrm{C}^+[0,1]$, then (2.6) is equivalent to

$$\Pr\left(\frac{1}{n}\max_{1\leq i\leq n}Y_i(s)\in\mathcal{B}\right)\to\Pr(\eta(s)\in\mathcal{B})=G(\mathcal{B}),\quad n\to\infty,$$

where G represents the unit Fréchet distribution. Let $F(\mathcal{B})=\Pr(Y(s)\in\mathcal{B})$, then we have

$$\Pr\left(\frac{1}{n}\max_{1\leq i\leq n}Y_i(s)\in\mathcal{B}\right)=F^n(n\mathcal{B})$$

and thus

$$n(1 - F(n\mathcal{B})) \to -\log G(\mathcal{B}).$$

Define the sequence of measures ν_n as

$$\begin{aligned}\nu_n(\mathcal{B}^c) &= n(1 - F(n\mathcal{B})) \\ &= n\Pr(Y \notin n\mathcal{B}) \\ &= n\Pr(n^{-1}Y \in \mathcal{B}^c),\end{aligned}$$

hence we have $\nu_n(\mathcal{B}) = n\Pr\{n^{-1}Y \in \mathcal{B}\}$, where $\mathcal{B}$ is a Borel set contained in $\mathrm{C}^+[0,1]$. Note that the measure ν_n verifies the sub-additivity property and thus is well defined (Schilling, 2005, Proposition 4.3, p.23).

Let $\bar{C}^+_\varrho[0,1] = (0,\infty] \times \bar{C}^+_1[0,1]$, where $(0,\infty]$ is equipped with the metric $\varrho(x,y) := |1/x - 1/y|$ and $\bar{C}^+_1[0,1]$ represents the space of positive functions in $\mathrm{C}[0,1]$ with supremum norm equal to 1.

Theorem 2.2 (de Haan and Ferreira (2006, Theorem 9.3.1))**.** *Let $Y, Y_1, Y_2, \ldots$ be i.i.d. stochastic processes in $\mathrm{C}^+[0,1]$, if (2.6) holds, then*

$$\nu_n \xrightarrow{\mathcal{D}} \nu$$

in $\bar{C}^+_\varrho[0,1]$. Equivalently, for every Borel set $\mathcal{B}$ in the space of positive functions of $\mathrm{C}[0,1]$ such that $\inf\{\|f\|_\infty : f \in \mathcal{B}\} > 0$ and $\nu(\partial\mathcal{B}) = 0$, where $\partial\mathcal{B}$ denotes the boundary of the set $\mathcal{B}$, then we have

$$\lim_{n\to\infty} \nu_n(\mathcal{B}) = \nu(\mathcal{B}).$$

Moreover the relation between the probability distribution of η and the measure ν is that for $m = 1, 2, \ldots,$

$$Pr\{\eta \in \mathrm{A}_{\boldsymbol{K},\boldsymbol{x}}\} = \exp\left\{-\nu\left(\mathrm{A}^c_{\boldsymbol{K},\boldsymbol{x}}\right)\right\} \tag{2.7}$$

with, for $\boldsymbol{K} = (K_1, \ldots, K_m)$ compact sets and $\boldsymbol{x} = (x_1, \ldots, x_m)$,

$$\mathrm{A}_{\boldsymbol{K},\boldsymbol{x}} := \left\{f \in \bar{C}^+_\varrho[0,1]; f(t) < x_i \text{ for } t \in K_i, i = 1, 2, \ldots, m\right\}.$$

Note that this weak convergence result is established in $\bar{C}^+_\varrho[0,1]$, a complete and separable metric space (CSMS), in which the space formed by the sequence of measures ν_n is relatively compact (see Daley and Vere-Jones (1988)).

Identically as in multivariate case (see Section 1.2.1, Chapter 1), the measure ν is called the exponent measure of the simple max-stable process which has the equivalent homogeneity property as (1.6). For any Borel set $\mathcal{B}$ in $\{f \in \mathrm{C}[0,1] : f \geq 0\}$ such that $\inf\{\|f\|_\infty : f \in \mathcal{B}\} > 0$

and $\nu(\partial\mathcal{B}) = 0$, and any $a > 0$

$$\nu(a\mathcal{B}) = a^{-1}\nu(\mathcal{B}). \tag{2.8}$$

Now, applying a one-to-one transformation to polar coordinates (equivalent to the transformation (1.7) in the multivariate case), $Q : \bar{C}^+[0,1] \to (0,\infty] \times \bar{C}_1^+[0,1]$, defined by

$$Q(f) = (\|f\|_\infty, f/\|f\|_\infty),$$

yields a spectral measure given as follows. Let $\mathcal{W}$ be a Borel set in $\bar{C}_1^+[0,1]$ and for $r > 0$ define the Borel set $\mathrm{B}_{r,\mathcal{W}} \subset \bar{C}^+[0,1]$ by

$$\mathrm{B}_{r,\mathcal{W}} := (r,\infty] \times \mathcal{W},$$

and thus we have the relationship

$$\mathrm{B}_{r,\mathcal{W}} = r(1,\infty] \times \mathcal{W} := r\mathrm{B}_{1,\mathcal{W}}.$$

The measure induced by the transformation Q is the product measure $\psi := \nu * Q$, i.e. $\psi(\mathrm{B}_{r,\mathcal{W}}) = \eta\{Q^{\leftarrow}(\mathrm{B}_{r,\mathcal{W}})\}$ and using the homogeneity property (2.8) we get

$$\begin{aligned}\psi(\mathrm{B}_{r,\mathcal{W}}) &= \nu\{f \in \bar{C}^+[0,1] : \|f\| > r, f/\|f\| \in \mathcal{W}\} \\ &= \nu\{rg \in \bar{C}^+[0,1] : \|g\| > 1, g/\|g\| \in \mathcal{W}\} \\ &= r^{-1}H'(\mathrm{B}_{1,\mathcal{W}})\end{aligned} \tag{2.9}$$

where r^{-1} represents the radial measure and $H'(\mathrm{B}_{1,\mathcal{W}}) := \nu\{g \in \bar{C}^+[0,1] : \|g\| > 1, g/\|g\| \in \mathcal{W}\}$ is a measure defined for each Borel set $\mathcal{W}$ in $\bar{C}_1^+[0,1]$, called the spectral measure.

This spectral decomposition provides the basis structure with which different representations of simple max-stable processes can be derived. There exist two main characterizations of such processes and both are detailed in the next section.

2.4 Characterization of max-stable processes

In this section we still focus on the simple max-stable process given by (2.6), and use the spectral decomposition of its dependence structure derived in the previous section. Two canonical representations of such processes are introduced with the aim of facilitating their construction. The parameter set $[0,1]$ is no longer considered here and replaced by any compact set Euclidean space: $\mathbb{S} \subset \mathbb{R}^d$. Until now the choice of $[0,1]$ was purely for convenience and it is easy to see that the set of interest is not essential. Thus, in order to be the most general, the following results are given in the set $\mathbb{S}$. Advantages and disadvantages of both representations are discussed at the end of this section.

A first characterization comes from the spectral representation of de Haan (1984) (see also Giné et al. (1990); Penrose (1992)) given in the following theorem

Theorem 2.3. *Let* $\{(\zeta_i, \psi_i), i \in \mathbb{N}\}$ *denote the points of a Poisson point process on* $(0, \infty) \times C^+(\mathbb{S})$*, where* $\psi_i, i \in \mathbb{N}$*, are sequences of non-negative continuous functions on* $\mathbb{S}$*, with intensity* $\zeta^{-2}\mathrm{d}\zeta\rho(\mathrm{d}\psi)$ *for some locally finite measure* ρ *such that*

$$\int \psi(s)\rho(\mathrm{d}\psi) = 1, s \in \mathbb{S}.$$

The max-stable process $\{\eta(s)\}_{s\in\mathbb{S}}$*, in* (2.6)*, has the same distribution as*

$$\left\{\max_{1\leq i} \zeta_i\psi_i(s)\right\}_{s\in\mathbb{S}}. \tag{2.10}$$

Remark 2.4 (de Haan (1984)). The above representation can also be expressed using a family of deterministic functions acting on a homogeneous Poisson point process on $(0, \infty) \times \mathbb{T}, \mathbb{T} \subset \mathbb{R}^d$. Specifically, there exist non-negative measurable functions $f_s(t)$ where $s \in \mathbb{S}$ and $t \in \mathbb{T}$ with

(i) for each $t \in \mathbb{T}$, $f_s(t)$ is continuous in s,

(ii) for each $s \in \mathbb{S}$,
$$\int_{\mathbb{T}} f_s(t)\mathrm{d}t = 1.$$

(iii)
$$\int_{\mathbb{T}} \sup_{s\in\mathbb{S}} f_s(t)\mathrm{d}t < \infty$$

such that the max-stable process $\{\eta(s)\}_{s\in\mathbb{S}}$, in (2.6), has the same distribution as

$$\left\{\max_{1\leq i} \zeta_i f_s(T_i)\right\}_{s\in\mathbb{S}} \tag{2.11}$$

where $\{(\zeta_i, T_i)\}_{1\leq i}$ denotes the points of a Poisson point process on $(0, \infty) \times \mathbb{T}$ with intensity measure $\zeta^{-2}\mathrm{d}\zeta \times \mathrm{d}\rho(t)$. The function f is often called the spectral function.

The ζ_i component can be interpreted as the intensity of an event i centred at T_i whose shape at s is given by $f_s(T_i)$. Furthermore, at any location $s \in \mathbb{S}$, the aggregated effects of all events centred a different locations t in $\mathbb{T}$ must be equal to one (condition (ii)). It is worth noting that taking $\mathbb{S} = \mathbb{T}$ and selecting a family of functions of the form $\{f(s - t) : s, t \in \mathbb{S}\}$ where f is an appropriate probability density function, would be convenient and easily interpretable (see for example Smith (1990)). Moreover the above spectral representation is not unique: different measure ρ might lead to the same max-stable process (de Haan and Ferreira, 2006, Remark 9.6.2). In applications, a choice of the spectral function has to be made. An example of the impact of a wrong choice is established in Kabluchko et al. (2009) and a criteria for the choice of spectral function is first investigated by Oesting et al. (2013).

A second spectral characterization of max-stable processes was given later by Schlather (2002), motivated by a rainfall example which will be detailed later on. The aim of this new representation is to be able to asses a broader range of extremal events. It is summarized in the theorem below.

Theorem 2.5 (Schlather (2002, Theorem 2)). *Let $\{\zeta_i\}_{1\leq i}$ be the points of a Poisson process on $(0,\infty)$ with intensity $\zeta^{-2}d\zeta$. Suppose that there exists a stochastic process W with continuous sample paths such that $\mathbb{E}[W^+(s)] = 1$ for all $s \in \mathbb{S}$ where W^+ denote the positive part of W, that is $W^+(s) = \max(0, W(s)), s \in \mathbb{S}$. Then the max-stable process $\{\eta(s)\}_{s\in\mathbb{S}}$ has the same distribution as*

$$\left\{\max_{1\leq i} \zeta_i W_i^+(s)\right\}_{s\in\mathbb{S}}, \tag{2.12}$$

where W_i are independent copies of W.

Note that, in (2.12), if we define the shape of the i-th event by $W_i^+(s) = f_s(T_i)$ then we recover (2.11). Furthermore, according to Lemma 3 from Schlather (2002), one may take $\zeta_i = 1/(E_1 + E_2 + \cdots + E_i)$ where $\{E_i\}_{1\leq i}$ are i.i.d. standard exponential random variables, independent of $\{W_i\}_{1\leq i}$. The following lemma provides the finite dimensional distributions of max-stable processes and shows that both representations agree with each other.

Remark 2.6. Stoev and Taqqu (2005) introduced the concept of extremal stochastic integrals producing a more general representation of max- stable processes (see also Wang and Stoev (2010)).

Lemma 2.7. *Let $s_1, \ldots, s_k \in \mathbb{S}$ and $x(s_1), \ldots, x(s_k) = x_1, \ldots, x_k > 0$, the finite dimensional distributions of a max-stable process η derived from Schlather's characterization* (2.12) *are*

$$Pr(\eta(s_1) \leq x_1, \ldots, \eta(s_k) \leq x_k) = \exp\left\{-\mathbb{E}\left[\max_{j\leq k} \frac{W^+(s_j)}{x_j}\right]\right\},$$

whereas using de Haan's characterization (2.11)

$$Pr(\eta(s_1) \leq x_1, \ldots, \eta(s_k) \leq x_k) = \exp\left\{-\mathbb{E}\left[\max_{j\leq k} \frac{f_{s_j}(T_j)}{x_j}\right]\right\}.$$

Proof. In the context of Theorem 2.5, let P denote the probability measure of W^+, we have:

$$\begin{aligned}
&\Pr(\eta(s_1) \leq x_1, \ldots, \eta(s_k) \leq x_k) \\
&\quad = \Pr\left(\max_{1\leq i} \zeta_i W_i^+(s_1) \leq x_1, \ldots, \max_{1\leq 1} \zeta_i W_i^+(s_k) \leq x_k\right) \\
&\quad = \Pr\left(\zeta_i W_i^+(s_1) \leq x_1, \ldots, \zeta_i W_i^+(s_k) \leq x_k;\ \text{for } i = 1, 2, \ldots\right) \\
&\quad = \Pr\left(\zeta_i \leq \min_{j\leq k} x_j / W_i^+(s_j);\ \text{for } j = 1, \ldots, k \text{ and } i = 1, 2, \ldots\right)
\end{aligned}$$

This corresponds to the probability of the event that no points of the Poisson process are above

the graph of $\min_{j\leq k} x_j/W^+(s_j)$. Hence this is equal to

$$\begin{aligned}
&\left(\int_{\mathbb{S}}\int_{\zeta>\min_{j\leq k} x_j/W^+(s_j)} \zeta^{-2}\mathrm{d}\zeta\mathrm{d}P\right)^0/0!\times\exp\left\{-\int_{\mathbb{S}}\int_{\zeta>\min_{j\leq k} x_j/W^+(s_j)} \zeta^{-2}\mathrm{d}\zeta\mathrm{d}P\right\}\\
&\quad=\exp\left\{-\int_{\mathbb{S}}\max_{j\leq k}\frac{W^+(s_j)}{x_j}\mathrm{d}P\right\}\\
&\quad=\exp\left\{-\mathbb{E}\left[\max_{j\leq k}\frac{W^+(s_j)}{x_j}\right]\right\}.
\end{aligned}$$

Secondly, the argumentation is identical in the context of Theorem 2.3. Consider the set

$$B=\{(\zeta,t):\zeta f_s(t)>x(s),\ \text{for at least one } s\in\mathbb{S}\}.$$

The event $\{\eta(s_1)\leq x_1,\ldots,\eta(s_k)\leq x_k\}$ occurs if and only if there are no points of the Poisson process that lie in B. The measure of B is

$$\int_{\mathbb{T}}\int_{\zeta>\min_{j\leq k} x_j/f_{s_j}(t)}\zeta^{-2}\mathrm{d}\zeta\mathrm{d}\rho(t)=-\int_{\mathbb{T}}\max_{j\leq k}\frac{f_{s_j}(t)}{x_j}\mathrm{d}\rho(t).$$

Hence, $\Pr(\eta(s_1)\leq x_1,\ldots,\eta(s_k)\leq x_k)=\exp\left\{-\int_{\mathbb{T}}\max_{j\leq k}\frac{f_{s_j}(t)}{x_j}\mathrm{d}\rho(t)\right\}$, which completes the proof. □

From Lemma 2.7 it is easy to see that, under both representations, the marginal distributions are unit Fréchet. The characterization of de Haan by (2.11) has the nice advantage of allowing an easy interpretation through storms. This will be developed in the next section (see Smith (1990)). Schlather's representation (2.12) doesn't allow such a nice and direct interpretation but we showed that it contains de Haan's characterization. Possible variations in the shape of the event do exist using Schlather's representation as the function f in (2.11) is replaced by a stochastic process.

In Chapter 1 the spectral measure was presented as a natural object to use when handling max-stable random vectors. However it is not as straightforward when dealing with stochastic processes as the dimension of the spaces on which the spectral measures are defined might not be the same. In this sense, using spectral functions (i.e. using the characterization of Theorem 2.3) might be easier as they are all defined on the same space.

According to representation (2.10), the max-stable process η is constructed as the pointwise maximum of an infinite number of points. Exact simulation of the process is then in general not straightforward. Nevertheless, not all the points (ζ,ψ) are necessary for the construction of η since only the maximum over all functions $\zeta\psi$ counts. de Haan and Ferreira (2006, Corollary 9.4.4) demonstrate that, under mild conditions, only a finite number of points (ζ,ψ) is required to evaluate the process at a point $s\in\mathbb{S}$. Furthermore Schlather (2002) and Oesting et al. (2013) provided stopping rules which led to exact simulations.

Being able to simulate max-stable processes is important. For example, when in high-

dimensions (or dimensions higher than 2), the marginal distributions of the process might not have a close form and is then addressed by simulations. Simulations are also used in applications when some characteristics of a parametric model cannot be calculated (Buishand et al. (2008) and Blanchet and Davison (2011)).

The emphasis on the key role of simulations is also put when trying to do predictions at unobserved locations. This is done estimating the distribution at a location of interest (unobserved) conditionally to the observations. This requires simulation procedures commonly called conditional simulations (see Dombry and Eyi-Minko (2013) and Dombry et al. (2013)).

Parametric models arise from the characterizations (2.10) and (2.12). The mostly widely used models are listed chronologically and their features (such as their distributions) are described in Section 2.6.

2.5 Summary measures for extremal dependence

When modelling extremes of a stochastic process, information about the dependence is essential, especially as this summarizes the evolution of the dependence through space or time. Quite often the study area is large enough to make the assumption that the marginal distribution varies across the area and characterizes the large scale effects. Small scale effects are due to the spatial nature of individual events and their influence on multiple locations, this corresponds to the dependence structure.

This section relies on the dependence measure given in Section 2.3. The dependence structure between the components of a max-stable random vector is either provided through the exponent measure ν or through its spectral decomposition (2.9), i.e. by the spectral measure H'. However, simpler and easier interpretable summary measures of the dependence are needed.

One of the main metric that captures dependence is the well-known *extremal coefficient function* which appeared in Smith (1990) and was later developed in Schlather and Tawn (2003). It consists of an extension of the definition of the extremal coefficient introduced in the multivariate case and given in (1.13) and now depends on the separation between two parameter values where the process is evaluated. Recall that for all $s_1, \dots, s_k \in \mathbb{R}^d$ and $x_1, \dots, x_k \in \mathbb{R}_+$, the finite-dimensional distribution function is

$$\begin{aligned}\Pr\left(\eta(s_1) \leq x_1(s_1), \dots, \eta(s_k) \leq x_k(s_k)\right) &= \exp\left\{-\int_{\mathbb{T}} \max_{j\leq k} \frac{f_{s_j}(t)}{x_j(s_j)} \mathrm{d}\nu(t)\right\} \\ &= \exp\{-V(x_1(s_1), \dots, x_k(s_k)\}.\end{aligned}$$

Considering that $\eta(\cdot)$ is stationary, $h \in \mathbb{R}^d$ and using the homogeneity property of the exponent measure (2.8), for all $x \in \mathbb{R}_+$, we can write

$$\Pr\{\eta(s_1) \leq x, \dots, \eta(s_k) \leq x\} = \exp\{-V(1, \dots, 1)/x\} = \exp\{-\vartheta(s_1, \dots, s_k)/x\},$$

where

$$\vartheta(s_1, \ldots, s_k) = \int_{\mathbb{T}} \max_{j \leq k} \{f_{s_j}(t)\} \mathrm{d}\nu(t), \quad \forall h \in \mathbb{R}^d.$$

Erhardt and Smith (2012) showed the use of high-order extremal coefficient can be meaningful and recently Genton et al. (2015) highlighted the possibility of working with an extended version of the extremal coefficient when focusing on multivariate spatial extremes. However, the pairwise extremal coefficient function is often taken into account in the literature and can be considered as a function of the separation h between two parameter values. It then reduces to

$$\Pr\{\eta(s) \leq x, \eta(s+h) \leq x\} = \exp\{-V(1,1)\} = \exp\{-\vartheta(h)/x\},$$

where

$$\vartheta(h) = \int_{\mathbb{T}} \max\{f_s(t), f_{s+h}(t)\} \mathrm{d}\nu(t).$$

This pairwise coefficient ranges between 1 (complete dependence) and 2 (independence). Properties of this extremal function are listed in Schlather and Tawn (2003, Theorem 3). While the extremal coefficient represents the dependence of the process at two different parameter values where at least one component exceeds a high threshold, considering the upper tail dependence, previously introduced in the multivariate case by (1.16), we can represent the dependence of the process at two different parameter values where both components exceed a high threshold.

It is straightforward to prove the following relationship

$$\chi(h) = 2 - \vartheta(h).$$

Other summaries of the spatial dependence are present in the literature. In geostatistics, the study of the spatial dependence structure of a process, generally Gaussian, is assessed by its semi-variogram

$$\gamma(s_1 - s_2) = \frac{1}{2}\mathbb{E}[(Z(s_1) - Z(s_2))^2], s_1, s_2 \in \mathbb{S}. \tag{2.13}$$

However it requires some assumptions (for instance finiteness of the mean), which might not always be met when focusing on extremes (for example for unit Fréchet distributions). Hence some variogram based approaches have recently emerged to bypass these assumption barriers. An important work based on madograms was done by Cooley et al. (2006). The drawbacks of madograms are identical to the ones for variograms, hence Cooley et al. (2006) started first by considering GEV distributions with shape parameter less than one in order to ensure the finiteness of first and second order moments. The choice of the GEV parameters being arbitrary, a second summary was established by replacing the value of the max-stable random field in the madogram approach by a function of it. This function being the unit Fréchet distribution this new summary is called the *F-madogram* and it exhibits clear links with the extremal coefficient. Based on the fact that this last madogram approach doesn't characterize the full dependence of a random field, Naveau et al. (2009) introduce the *λ-madogram*.

2.6 Max-stable models

Based on (2.10) and (2.12), some parametric models for spatial extremes are created. Here we list of the most encountered models in the literature, provide their finite dimensional distribution in dimension 2 or higher and provide an interpretation when possible. The dependence structure of each and every of these models will be summarized by their extremal coefficient and also their angular density; examples will be provided.

2.6.1 The Smith model

First model to appear in the seminal unpublished technical report from Smith (1990), this model's foundations lay on de Haan's spectral representation (2.11) and has the big advantage to allow an easy interpretation as a storm model. Here $\mathbb{T}$ denotes the space of storm centers and the measure ν represents the distribution of the storms over $\mathbb{T}$. The shape of a given storm centered at t_i is given by $f(\cdot, t_i)$ and its magnitude is ζ_i. Then $\zeta_i f(s, t_i)$ can be interpreted as the amount of rainfall received at location $s \in \mathbb{S}$ for a storm of magnitude ζ_i and centered at t_i. The process $\eta(s)$ denotes the maximum rainfall received at location s over an infinite number of independent storms.

Following the suggestion from the previous section which says that a convenient choice is to consider where $\mathbb{S} = \mathbb{T}$ and f an appropriate density function. Smith takes $\mathbb{S} = \mathbb{T} = \mathbb{R}^d$, f a d-variate normal probability density function with mean 0 and covariance matrix Σ and lets ν to the Lebesgue measure.

This model is often called the *Gaussian extreme value model* or simply the *Smith model*. If we consider k parameter values $\{s_1, \ldots, s_k\}$ in $\mathbb{S} = \mathbb{R}^d$, then the k-dimensional cdf of $(\eta(s_1), \ldots, \eta(s_k))$ at $X = (x_1, \ldots, x_k)^\top \in \mathbb{R}_+^k$, $x_i \equiv x_i(s_i) \in \mathbb{R}_+, \forall i$, is given by

$$\Pr\left(\eta(s_1) < x_1, \ldots, \eta(s_k) < x_k\right) = \exp\left\{-\sum_{j=1}^{k} x_j^{-1} \Phi_{k-1}(q^{(j)}(x); \Sigma^{(j)})\right\}, \tag{2.14}$$

where Φ_{k-1} is the $k-1$ dimensional normal cdf and $q^{(j)}(x) = \left(q_1^{(j)}(x), \ldots, q_{j-1}^{(j)}(x), q_{j+1}^{(j)}(x), \ldots, q_k^{(j)}(x)\right)^\top \in \mathbb{R}^{k-1}$, $q_p^{(j)}(x) = q_p^{(j)}/2 - \log(x_j/x_p)$ for $j, p = 1, \ldots, k$ and $q_p^{(j)} = (s_j - s_p)^\top \Sigma^{-1}(s_j - s_p)$. Moreover Σ is a $d \times d$ positive definite matrix and $\Sigma^{(j)} = \left(s_j \mathbb{1}_{k-1}^\top - S_{-j}\right)^\top \Sigma^{-1} \left(s_j \mathbb{1}_{k-1}^\top - S_{-j}\right)$, is a $(k-1) \times (k-1)$ matrix where $S = (s_1, \ldots, s_k) \in \mathbb{R}^{d \times k}$, $S_{-j} = S \backslash s_j \in \mathbb{R}^{d \times (k-1)}$ for $j = 1, \ldots, k$ and $\mathbb{1}_{k-1}$ is the $k-1$ dimensional indicator function.

The associated extremal coefficient function is

$$\vartheta(s_1, \ldots, s_k) = \sum_{j=1}^{k} \Phi_{k-1}(q^{(j)}/2; \Sigma^{(j)}),$$

where the components of $q^{(j)}$ are $q_p^{(j)}$ for $j, p = 1, \ldots, k$. The deduced pairwise extremal coeffi-

cient is

$$\vartheta(h) = \vartheta(s, s+h) = 2\Phi_1\left(\frac{\sqrt{h^\top \Sigma^{-1} h}}{2}\right), \quad h \in \mathbb{R}^d. \tag{2.15}$$

Furthermore, from the finite-dimensional distribution (2.14) the spectral density is derived by

$$h(\boldsymbol{w}; \boldsymbol{\theta}) = \phi_{k-1}(q^{(1)}(\boldsymbol{w}); \Sigma^{(1)}) \left\{ w_1^2 \prod_{j=2}^{k} w_j q_j^{(1)} \right\},$$

where ϕ_{k-1} represents the $k-1$ dimensional normal density. The expression of the angular density given above is equivalent to the angular density of the Hüsler-Reiss model given in Section 1.3.4 in the multivariate setting where $\lambda_{j,p}$ is replaced by $\sqrt{2q_p^{(j)}}$.

Different Smith models may differ by their covariance structure which can be isotropic or anisotropic. Isotropy implies that the covariance is a function of distance only. If Σ is a diagonal matrix then we are in the isotropic case. Moreover if its diagonal terms are all equal then the shape of the events (induced by η) are circular whereas if the diagonal terms are different then the shape is elliptic and provide the rates of increase along each axis. If Σ is not a diagonal matrix then we are in the anisotropic case and its eigen-decomposition is $\Sigma = RDR^\top$ where R is a rotation matrix and D a diagonal matrix whose diagonal elements are the corresponding eigen-values. The eigen-values provide the rates of increase along each axis and the rotation matrix the angle of the rotation.

Figure 2.1 represents realizations and the extremal coefficient (2.15) for the Smith model under both the isotropic and anisotropic scenarios. The process is defined for $s \in \mathbb{S} = [0,1]^2$ and some realizations are given on the top row (the margins are transformed to Gumbel margins for visualization purposes). The random fields were simulated through the `R` package `SpatialExtremes`, Ribatet et al. (2013). Three different covariance matrices $\Sigma = (\sigma_{12}, \sigma_{12}, \sigma_{22})$ are considered. For the left column, the covariance matrix is $\Sigma = (0.012, 0, 0.012)$ which consist in the isotropic case where the rate of increase along both axes are the same. Consequently the impact of the 'storms' given by the Smith model is circular and can be assessed by looking at the realization of the process or the extremal coefficient. Clearly, in this case the further away a location is to another the more independent they get no matter the direction. The middle and right columns exhibit anisotropy. The middle column has covariance matrix $\Sigma = (0.012, 0, 0.006)$ which means that an increase of two units along the x-axis leads to an increase of one unit along the y-axis. The shape of the dependence structure is then elliptic. Independence is reached twice as fast by moving along the y-axis than the x-axis. The right column adds a rotation components which occurs due to the presence of a non-zero correlation coefficient on the matrix $\Sigma = (0.012, 0.01, 0.012)$. Lastly, extensions of the Smith model to other densities such as Student-t and Laplace densities were given in de Haan and Pereira (2006).

2.6.2 The Schlather model

Schlather (2002) gave a new canonical representation of max-stable processes via Theorem

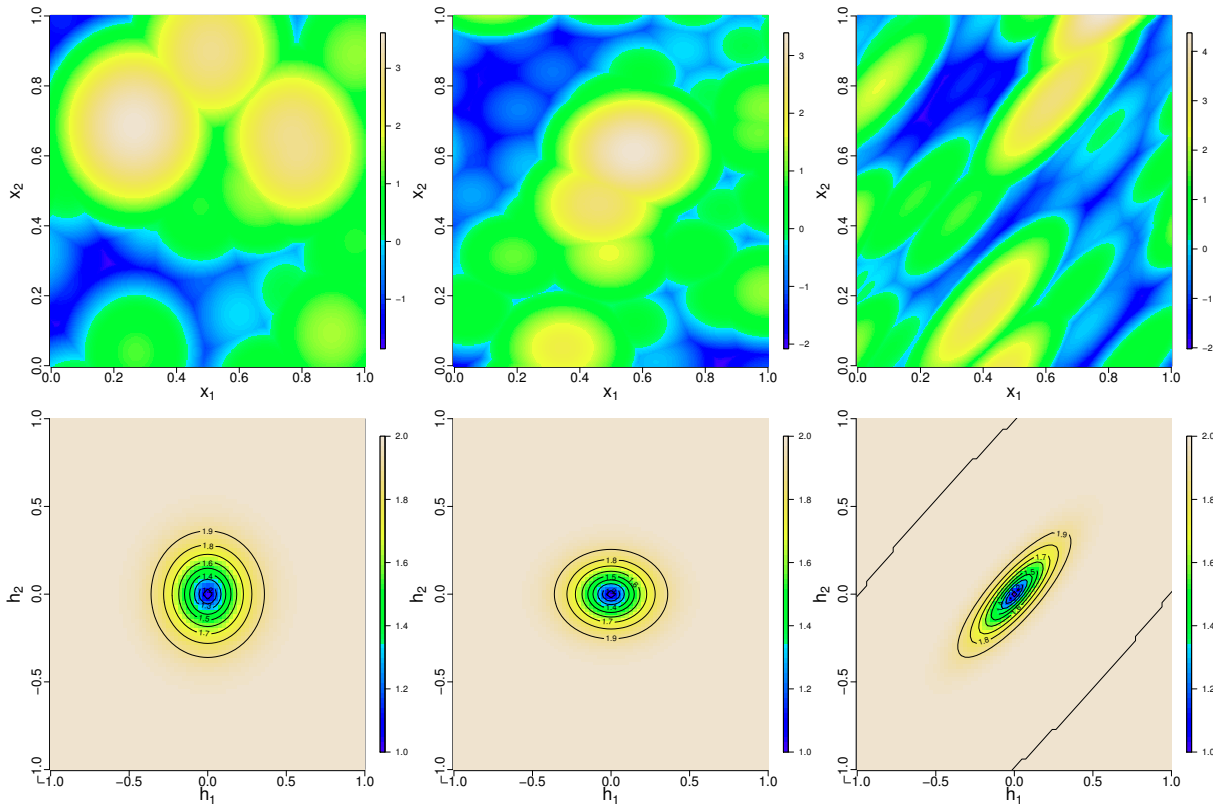

Figure 2.1: Smith model on $\mathbb{S} = [0,1]^2$ with different covariance matrices Σ. Top row represents a realization of the process and bottom row shows the extremal coefficient. The left column deal with the isotropic case with respectively $\sigma_{11} = \sigma_{22} = 0.012$ and $\sigma_{12} = 0$. The middle and right column show the anisotropic case. The middle column has covariance matrix whose components are $\sigma_{11} = 0.012$, $\sigma_{22} = 0.006$ and $\sigma_{12} = 0$, which leads to an elliptic form of the dependence. The right column has a dependence structure given by $\sigma_{11} = \sigma_{22} = 0.012$ and $\sigma_{12} = 0.01$, adding a rotation effect. The margins are transformed to be Gumbel.

2.5 motivated by the nature of some spatial extremal event whose specificities are different from the ones given by the Smith model. The existence of different types of precipitation was highlighted proving the need of different class of models. The Smith model introduced above suits events that are convective, i.e., centered and with high intensity in a particular location while there is almost no rainfall elsewhere. This new characterization puts the emphasis on a different class of precipitation events, called cyclonic, that are spread over a region with variable rainfall and clearly for which the shape of Gaussian densities is too restrictive. This leads to a new model named the extremal Gaussian process, shortly: the Schlather model.

Let W be a stationary Gaussian process on $\mathbb{S}$ with correlation function ρ. The expected value of the positive part of the process is $\mathbb{E}[W^+(s)] = (2\pi)^{-1/2}$ hence we need to add a $\sqrt{2\pi}$ term in order to fulfill the condition on the mean. The process is then given by

$$\{\eta(s)\}_{s\in\mathbb{S}} = \left\{\sqrt{2\pi}\max_{1\leq i}\zeta_i W_i^+(s)\right\}, \tag{2.16}$$

where W_i are i.i.d. replicates of W. Relating the model to precipitations, $\zeta_i W_i^+(\cdot)$ can be thought as the daily spatial rainfall events with the same spatial dependence structure but different magnitude. Following the Gaussian nature of the random field, this model is also sometimes referred as the extremal Gaussian model.

The bivariate dimensional distribution is given, for all $x_1, x_2 \in \mathbb{R}_+$, by

$$\Pr\left(\eta(s_1) < x_1, \eta(s_2) < x_2\right) = \exp\left\{-\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\left(1+\left[1-\frac{2\{\rho\left(h\right)+1\}x_1x_2}{(x_1+x_2)^2}\right]^{1/2}\right)\right\}, \quad (2.17)$$

where $h \in \mathbb{R}^d$. The general k-dimensional distribution function isn't available since, following the steps of the proof in (Schlather, 2002, Appendix), a differential equation needs to be solved but its solution is not known. However, the k-dimensional distribution and the spectral density are obtained as a special case of the extremal-t model with the number of degrees of freedom equal to one. The pairwise extremal coefficient function is derived from (2.17),

$$\vartheta(h) = \vartheta(s, s+h) = 1 + \sqrt{\frac{1-\rho(h)}{2}}, \quad h \in \mathbb{R}^d. \quad (2.18)$$

Numerous parametric models for the correlation structure ρ are available in the literature (see Chilès and Delfiner (1999); Gelfand et al. (2010) for a review). The main families of correlation functions are Whittle-Matérn, Cauchy, Power Exponential and Bessel which have a range and smoothness parameter. Other families are listed and available in the `R` package `RandomFields` (Schlather et al., 2015). Anisotropy can be obtained by introducing a rotation matrix.

Figure 2.2 considers the Power Exponential and Bessel correlation functions (left to right) and realizations of the process, extremal coefficients and correlation functions (top to bottom) are given. Generation of the random fields was performed using the `R` package `RandomFields` (Schlather et al., 2015). As expected from (2.18), the behaviour of the extremal coefficient is driven by the correlation function. The left and middle columns provide examples respectively of the isotropic and anisotropic Power Exponential correlation. For the anisotropic case a rotation of $\pi/4$ is considered and for an increase by 1 unit along the x-axis there is an increase by 3 units along the y-axis. When considering the Bessel correlation it can be shown that is has the property of being a non monotonic function and also allows negative values which is not the case of the Power Exponential. The impact of the correlation function on the process is clear, the stronger the slope of the function near the origin (short distance away from a location), the rougher the process gets. The rate of decay of the correlation is the slowest for the Bessel correlation resulting in a smoother field.

Suppose that $\rho(h) \to 0$ as $h \to \infty$ then $\lim_{h\to\infty} \vartheta(h) = 1 + 2^{-1/2}$. Furthermore if the correlation function on $\mathbb{S} = \mathbb{R}^2$ is positive definite isotropic then it can't take values lower than -0.403 which means that $\vartheta(h) \leq 1.838, \forall h \in \mathbb{S}$. These two assertions show that no matter the distance between two points, independence won't be reached (independence is reached when $\rho(h) = -1$). Note that all of the parametric families listed above don't reach independence as their range is $[0, 1]$ except for the Bessel correlation function as seen in Figure 2.2.

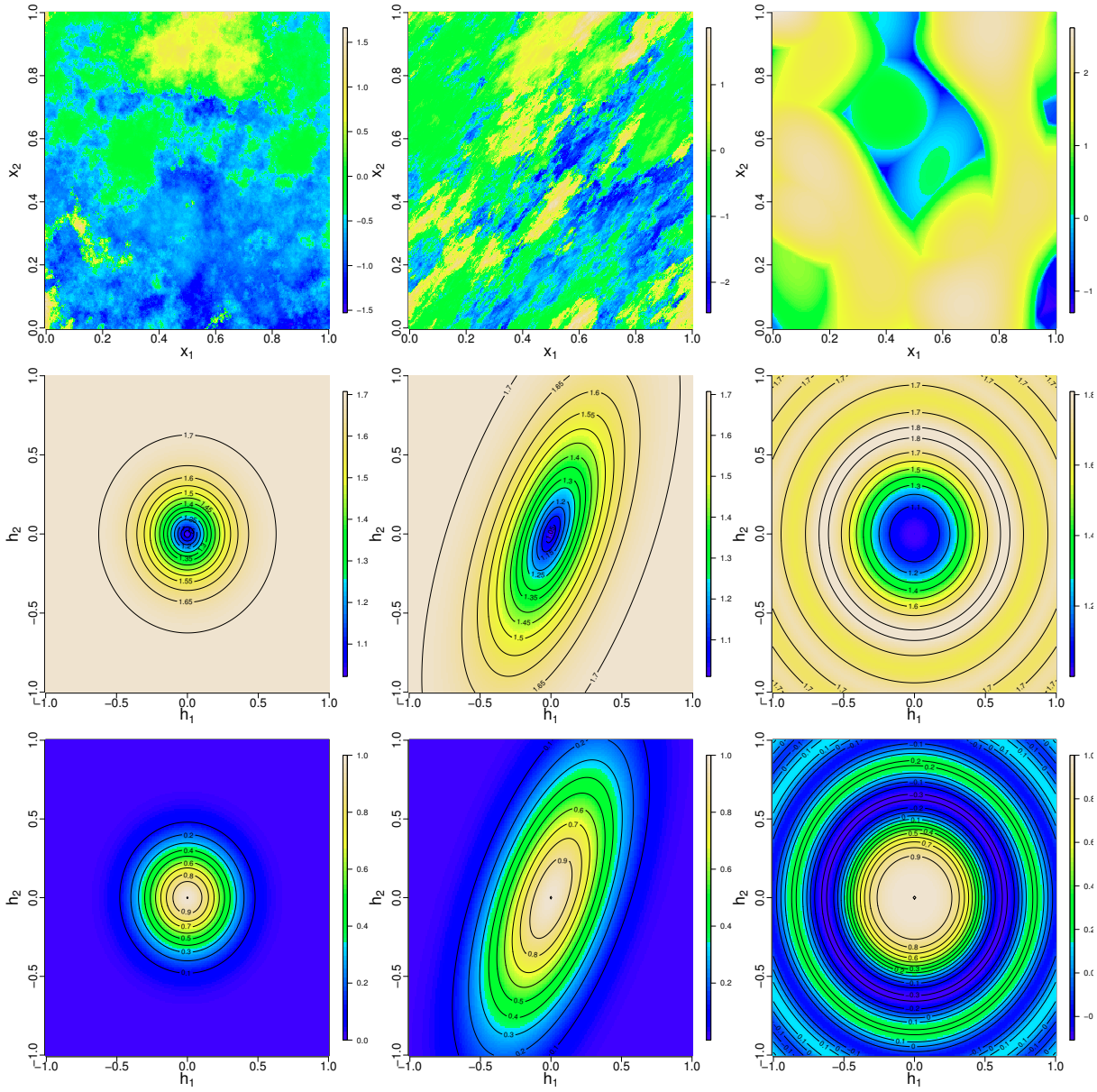

Figure 2.2: Representation of Schlather processes on $\mathbb{S} = [0,1]^2$. The top row represents some realizations of the process while the second and third row shows respectively the extremal coefficient and the correlation function. The first and second column consider isotropic and anisotropic power exponential correlation function $\rho(h), h \in [-1,1]^2$ with scale 0.1 and smoothness 1. The model represented on the right column has Bessel correlation function with scale 0.1 and smoothness 0.2. The margins are transformed to be Gumbel.

Davison and Gholamrezaee (2012) bypass this problem by replacing the scaled stationary standard Gaussian processes $W_i(s)$ by the stationary process

$$W_j^{\mathcal{B}}(s) = W_j(s)\mathbb{I}_{\mathcal{B}_j}(s - X_j),$$

where $\mathbb{I}_{\mathcal{B}}$ is the indicator function of a compact random set $\mathcal{B} \in \mathbb{S}$ with volume $|\mathcal{B}|$, $\{\mathcal{B}_j\}$ are

independent replicates of $\mathcal{B}$ and X_j are points of a Poisson process with rate $\mathbb{E}(|\mathcal{B}|)^{-1}$. This means that the daily spatial rainfall events characterize by $\zeta_j W_j^+(\cdot)$ are only affecting the region $\mathcal{B} + X_j$. For $x_1, x_2 \in \mathbb{R}_+$, the distribution function (2.17) then becomes

$$\Pr\left(\eta(s_1) < x_1, \eta(s_2) < x_2\right) = \exp\left\{-\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\left(1 - \frac{\alpha(h)}{2}\left[1 - \frac{2\{\rho\left(h\right)+1\}x_1x_2}{(x_1+x_2)^2}\right]^{1/2}\right)\right\},$$

where $\alpha(h) = \mathbb{E}[|\mathcal{B} \cap (h + \mathcal{B})|]/\mathbb{E}[|\mathcal{B}|]$ lies on the unit interval. It is now straightforward to see that the extremal coefficient given by

$$\vartheta(h) = 2 - \alpha(h)\left(1 - \sqrt{\frac{1-\rho(h)}{2}}\right), \quad h \in \mathbb{R}^d,$$

can reach complete dependence and independence. If the distance h is large enough then the set $\mathcal{B} \cap (\mathcal{B} + h)$ can be empty, leading to $\alpha(h) = 0$ and thus $\vartheta(h) = 2$. Flexible shapes of the set $\mathcal{B}$ can be used to produce simple forms of $\alpha(h)$; Davison and Gholamrezaee (2012) first consider a disc with fixed radius and then with random radius GPD distributed to be more realistic. Elliptical shapes can also be included.

2.6.3 The Brown-Resnick model

The Brown-Resnick model get its name from the seminal paper from Brown and Resnick (1977) and whose approach is based on the Schlather canonical representation (2.12) where the stochastic process W is given by

$$W(s) = \exp\{\epsilon(s) - \sigma^2(s)/2\}, s \in \mathbb{R}, \tag{2.19}$$

where $\epsilon(s)$ is a standard Brownian motion with variance $\text{Var}(\epsilon(s)) = \sigma^2(s) = |s|$. Moreover, even though $W(s)$ is not stationary the resulting max-stable process η is stationary. Kabluchko et al. (2009) extend this statement by proving that by subtracting an appropriate drift term to any Gaussian process ϵ with stationary increments leads to a stationary max-stable process. Because of the stationarity of the increments the law of W is completely characterized by the (semi)-variogram and the variance. The generalized Brown-Resinck process is then given by

$$\{\eta(s)\}_{s\in\mathbb{S}} = \left\{\max_{1\leq i} \zeta_i \exp\{\epsilon_i(s) - \gamma(s)\}\right\}_{s\in\mathbb{S}}, \tag{2.20}$$

where ϵ_i are i.i.d. replicates of a zero mean Gaussian process with stationary increments and semi-variogram $\gamma(h) = \text{Var}(\epsilon(s+h) - \epsilon(s))/2$. The finite dimensional distributions were given in Huser and Davison (2013) and have also simultaneously been derived and given in the next chapter. Again letting $x_i \equiv x_i(s_i) \in \mathbb{R}_+$, their expression is the following

$$\Pr\left(\eta(s_1) < x_1, \ldots, \eta(s_k) < x_k\right) = \exp\left\{-\sum_{j=1}^{k} x_j^{-1} \Phi_{k-1}(q^{(j)}(x); R^{(j)})\right\}, \tag{2.21}$$

where $q^{(j)}(x) \in \mathbb{R}^{k-1}$ whose elements are $q_p^{(j)}(x) = q_p^{(j)}/2 - \log(x_j/x_p)/q_p^{(j)}$ with $j, p = 1, \ldots, k$, $p \neq j$ and $q_p^{(j)} = \sqrt{2\gamma_{j,p}}$, $\gamma_{j,p} = \gamma(s_j - s_p)$. Moreover $R^{(j)} = (\gamma_{j,u} + \gamma_{j,v} - \gamma_{u,v})/\{2(\gamma_{j,u}\gamma_{j,v})^{1/2}\}$, for $u, v = 1, \ldots, k$ and $u, v \neq j$, is a $(k-1) \times (k-1)$ correlation matrix.

The associated extremal coefficient function is

$$\vartheta(s_1, \ldots, s_k) = \sum_{j=1}^{k} \Phi_{k-1}(q^{(j)}/2; R^{(j)}),$$

where the components of $q^{(j)}$ are $q_p^{(j)}$ for $j, p = 1, \ldots, k$, $p \neq j$. The deduced pairwise extremal coefficient is

$$\vartheta(h) = \vartheta(s, s+h) = 2\Phi_1\left(\sqrt{\frac{\gamma_{s,s+h}}{2}}\right), \quad h \in \mathbb{R}^d. \tag{2.22}$$

Based on the finite-dimensional distribution (2.21), the angular density of the model has the same form as the angular density of the Hüsler-Reiss model in Section 1.3.4 where the coefficients $\lambda_{i,j}$ are replaced by the variograms $\gamma_{i,j}$.

One of the most important Gaussian random process is the fractional Brownian motion (Mandelbrot and Van Ness, 1968) $\epsilon = \{\epsilon(s), s \in \mathbb{S}\}$ which is a centered Gaussian random process with covariance function

$$\mathbb{E}[\epsilon_i(s_1)\epsilon_j(s_2)] = \frac{\delta_{ij}}{2}\left(\|s_1\|^\alpha + \|s_2\|^\alpha - \|s_1 - s_2\|^\alpha\right), \forall s_1, s_2 \in \mathbb{S},$$

where $0 < \alpha \leq 2$, $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and $\|\cdot\|$ is the Euclidean norm in $\mathbb{S}$. Note that the parameter α can be thought as the self-similarity parameter of ϵ. Also, the semi-variogram is easily calculated and is $\gamma(h) \equiv \gamma(s, s+h) = \|h\|^\alpha, h \in \mathbb{S}$. This class of random processes also has the big advantage of being isotropic.

In the case $\alpha = 1$ we recover the standard Brownian motion case and the model reduces to the one introduced by Brown and Resnick (1977) with (2.19). When $\alpha = 2$ then $\epsilon(s) = s^\top \epsilon(\mathbf{1})$, $\epsilon(\mathbf{1})$ being multivariate standard Gaussian random variable (see Samorodnitsky and Taqqu (1994, pp.318–319)), hence we can write $\epsilon(s) = s^\top \Sigma^{-1/2} Z$, where Z is a multivariate random variable in $\mathbb{S}$ with variance-covariance matrix Σ. Consequently, $\gamma(h) = \frac{1}{2} h^\top \Sigma^{-1} h, h \in \mathbb{S}$, which corresponds to the Smith model from Section 2.6.1. Additionally, if the L_1 norm is inserted in the expression of the variogram, the resulting process is the one considered in Buishand et al. (2008) and de Haan and Zhou (2008).

A strong interest on the simulation of Brown-Resnick processes has been shown in the recent years. An accurate simulation of such processes is complicated: the naive simulation approach by finite approximation may fail, large approximation errors occurs on large intervals, calculations are computationally demanding; see for example Oesting et al. (2012); Oesting and Schlather (2014) and Dombry and Eyi-Minko (2013). Recently Dieker and Mikosch (2015) proposed a new representation of Brown-Resnick processes leading to the same finite-dimensional distributions and the ability to perform exact simulations.

According to the simulation results from Oesting et al. (2012), the realizations of the process

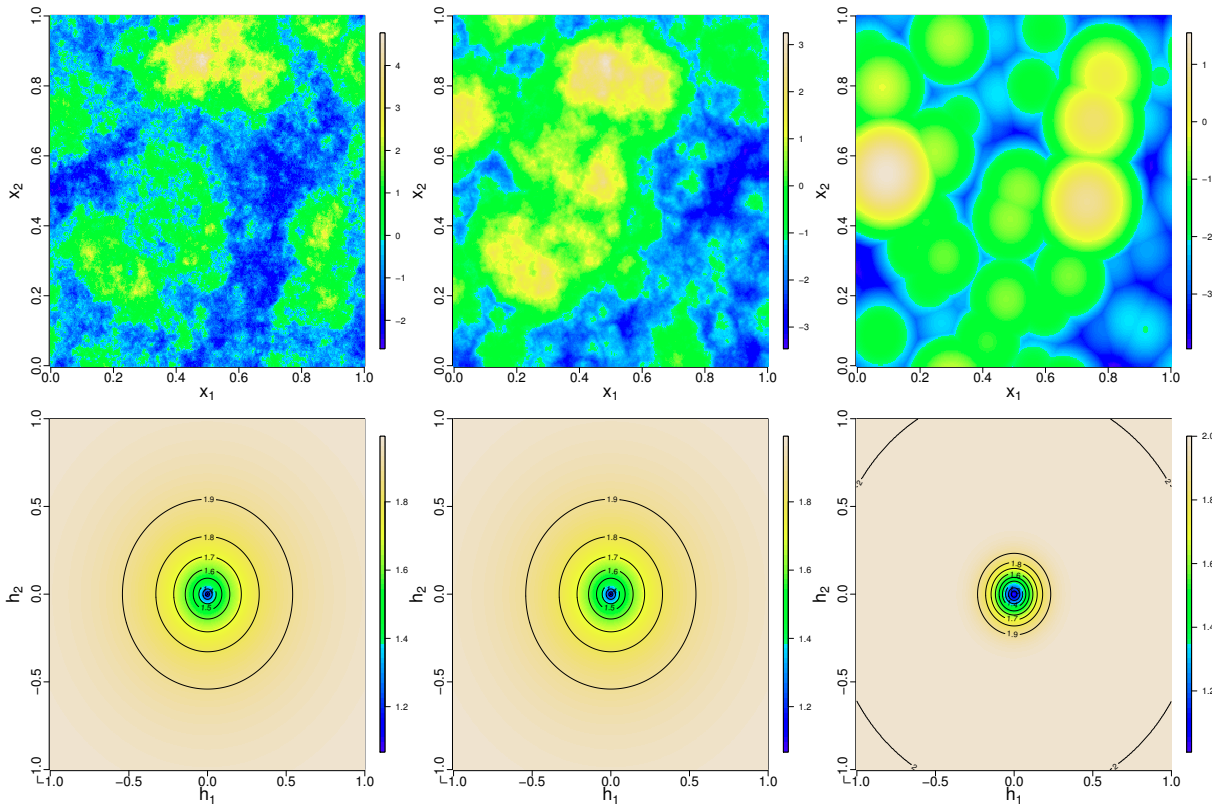

Figure 2.3: Representation of Brown-Resnick processes on $\mathbb{S} = [0,1]^2$ constructed by fractional Brownian motions i.e. with semi-variogram $\gamma(h) = \|h\|^\alpha, h \in \mathbb{S}, \alpha \in (0,2]$. Top row represents realizations of the process and bottom row shows the associated extremal coefficients. The left column deals with the standard Brownian case, $\alpha = 1$ given by (2.19). The middle considers the case $\alpha = 1.5$ whereas the right column corresponds to the isotropic Smith model ($\alpha = 2$). The margins are on the Gumbel scale.

given in the top row of Figure 2.3 have been simulated using the random shifts method available in the `R` package `RandomFields` (see Schlather et al. (2015) for details). Figure 2.3 shows the influence of the variogram on the smoothness of the max-stable process. From left to right the smoothness parameter α is increasing, taking the values 1, 1.5 and 2 until the isotropic case of the Smith model is reached. When $\alpha = 1$ this corresponds to the original process (2.19) introduced by Brown and Resnick (1977). The bottom row of the figure outlines the changes in the extremal coefficient as α increases. Clearly when $h \to 0$ then $\gamma(h) \to 0$ complete dependence is reached whereas if $\gamma(h)$ is unbounded then $\vartheta(h) \to 2$ as $\|h\| \to \infty$.

The generalized Brown-Resnick model given by (2.20) contains also the geometric Gaussian model given in Davison et al. (2012) where

$$W(s) = \exp\{\sigma\epsilon(s) - \sigma^2/2\}, s \in \mathbb{S},$$

where $\epsilon(s)$ is a standard Gaussian process with correlation function $\rho(h)$ and $\sigma(s) = \sigma, \forall s \in \mathbb{S}$. The form of its distribution function is identical to the ones of the Smith model (2.14) and of the

generalized Brown-Resnick (2.21) where the quantities $q_p^{(j)}$ would be respectively replaced by $2\sigma^2\{1-\rho_{j,p}(h)\}$ and $(2\sigma^2\{1-\rho_{j,p}(h)\})^{1/2}$. The difference is only due to the fact the Smith model's distribution specifies the covariance matrix whereas for the Brown-Resnick the (correlation) matrix. Thus the pairwise extremal coefficient can be deduced from (2.15) or (2.22) as

$$\vartheta(h) = 2\Phi\left(\sigma\sqrt{\frac{1-\rho(h)}{2}}\right), \quad h \in \mathbb{R}^d,$$

which reaches complete dependence as $\sigma \to 0$ or $\rho(h) \to 1$ and independence as $\sigma \to \infty$ for any ρ.

In Chapter 3, we will provide results on Brown-Resnick processes. Hence we will first provide a proof of the existence of such process (see Appendix B). Brown and Resnick (1977) have successfully shown that the pointwise maxima of Brownian motions and Ornstein-Ulhenbeck processes are converging to Poisson processes equivalent to the ones in (2.12). These have the stationarity property.

2.6.4 The extremal t model

A more flexible class of processes which generalizes the extremal Gaussian (2.16) has been introduced by Opitz (2013) and is commonly called the class of extremal-t processes. It is a generalization of the extreme value copula to infinite dimensions (see Demarta and McNeil (2005); Nikoloulopoulos et al. (2009); Davison et al. (2012); Ribatet and Sedki (2013)) which are proved to be more flexible than Hüsler-Reiss copulas that are themselves generalized by the Smith Gaussian model. Opitz (2013) defines the spectral representation of such processes based on the Schlather characterization of max-stable process (2.12) by taking

$$W_i^+(s) = c_\nu\{\max(\epsilon_i(s), 0)\}^\nu, \nu \geq 1$$

where $c_\nu = \sqrt{\pi}\nu^{1-\nu/2}/\Gamma((\nu+1)/2)$ and ϵ_i are i.i.d. replicates of a standard Gaussian process with partial correlation function ρ. Clearly as $c_1 = \sqrt{2\pi}$, the Schlather model arises as a special case when $\nu = 1$. The finite dimensional distributions are available and, letting $x_i \equiv x_i(s_i) \in \mathbb{R}_+$, defined as

$$\Pr\left(\eta(s_1) < x_1, \ldots, \eta(s_k) < x_k\right) = \exp\left\{-\sum_{j=1}^{k} x_j^{-1} T_{k-1}\left(q^{(j)}(x); R^{(j)}, \nu+1\right)\right\}, \tag{2.23}$$

where $T_k(\cdot; R, \nu)$ represents the k-dimensional Student-t cdf with correlation matrix R and degree of freedom ν, $q^{(j)}(x) \in \mathbb{R}^{k-1}$ whose elements are $q_p^{(j)}(x) = \sqrt{\frac{\nu+1}{1-\rho_{j,p}^2(h)}}\left(\left(\frac{x_p}{x_j}\right)^\nu - \rho_{j,p}(h)\right)$ for $j, p = 1, \ldots, k$, $p \neq j$. $R^{(j)}$ is a $(k-1)\times(k-1)$ partial correlation matrix defined as $R^{(j)} = \omega_{\bar{R}^{(j)}}^{-1}\bar{R}^{(j)}\omega_{\bar{R}^{(j)}}^{-1}$ with $\bar{R}^{(j)} = R_{-j,-j} - R_{-j,j}R_{j,-j}$ and $\omega_{\bar{R}^{(j)}} = \operatorname{diag}\left(\bar{R}^{(j)}\right)^{1/2}$ where the negative index means that the associated row or column has been removed. Considering the pairwise distribution and setting $\nu = 1$ yields the distribution function (2.17) of the Schlather model. The finite-

dimensional distribution of the Schlather model cannot be obtained through the construction (2.16) but it can be derived by the extremal-t distribution setting the degree of freedom equal to 1. conceptually this model is equivalent to the Extremal-t with degrees of freedom 1 and thus can be given by (2.23). The associated extremal coefficient function is

$$\vartheta(s_1, \ldots, s_k) = \sum_{j=1}^{k} T_{k-1}(q^{(j)}/2; R^{(j)}, \nu + 1),$$

where the components of $q^{(j)}$ are $q_p^{(j)} = \sqrt{\frac{\nu+1}{1-\rho_{j,p}^2(h)}}\,(1 - \rho_{j,p}(h))$ for $j, p = 1, \ldots, k$, $p \neq j$. The deduced pairwise extremal coefficient is

$$\vartheta(h) = \vartheta(s, s+h) = 2T_1\left(\sqrt{\frac{\nu+1}{1-\rho^2(h)}}(1 - \rho(h)); \nu + 1\right), \quad h \in \mathbb{R}^d. \tag{2.24}$$

If $\rho(h) = 0$, it is straightforward to see that the range of the extremal coefficient is $(1.5, 2)$: independence is reached as the degree of freedom gets large i.e. $\vartheta(h) \to 2$ as $\nu \to \infty$, whereas $\vartheta(h) \to 1.5$ as ν gets small. The degree of freedom ν is acting as a general dependence parameter.

The general k-dimensional angular density for the finite-dimensional distributions (2.23) remains identical to the one provided for the Extremal-t model in the multivariate context (Section 1.3.5).

Figure 2.4 displays, from top to bottom, realizations of the process (using `RandomFields R` package), extremal coefficients and correlation functions. The left and middle columns of Figure 2.4 display two Extremal-t models with same degree of freedom $\nu = 2$ but for the sake of being as complete as possible, with two correlation which differ from the ones presented in the Schlather model example. The left column introduces the widely used Whittle-Matérn correlation function with smoothness equal to 0.8 whereas the middle column displays the example of the damped cosine correlation function with smoothness 1 and scale 0.1. The effect of the scale on the correlation structure and the link between the correlation structure and the extremal coefficient are quite obvious. The scale being larger, the rate of decrease towards zero of the Whittle-Matérn correlation function is much slower than for the damped cosine with scale 0.1. As a result, according to the extremal coefficient the distance needed to reach independence is much larger (independence is actually never reached on $[0, 1]^2$ for the model with Whittle-Matérn correlation with such parameters). Comparing the middle and right columns highlights the general dependence role of the degree of freedom on the process itself. As ν increases independence is reached faster, leading to a rougher process. Finally it is worth noting that the bottom right panel gives a representation of the univariate damped cosine correlation. It shows, more explicitly than the bivariate representation in the bottom middle panel, the ability of this parametric model to take negative values.

To conclude, we have shown in this chapter that spatial extremes is an extensively studied area with complex challenges. An accurate understanding of the spatial dependence structure is fundamental and requires tools from Geostatistics. In Chapter 4 we define a more general

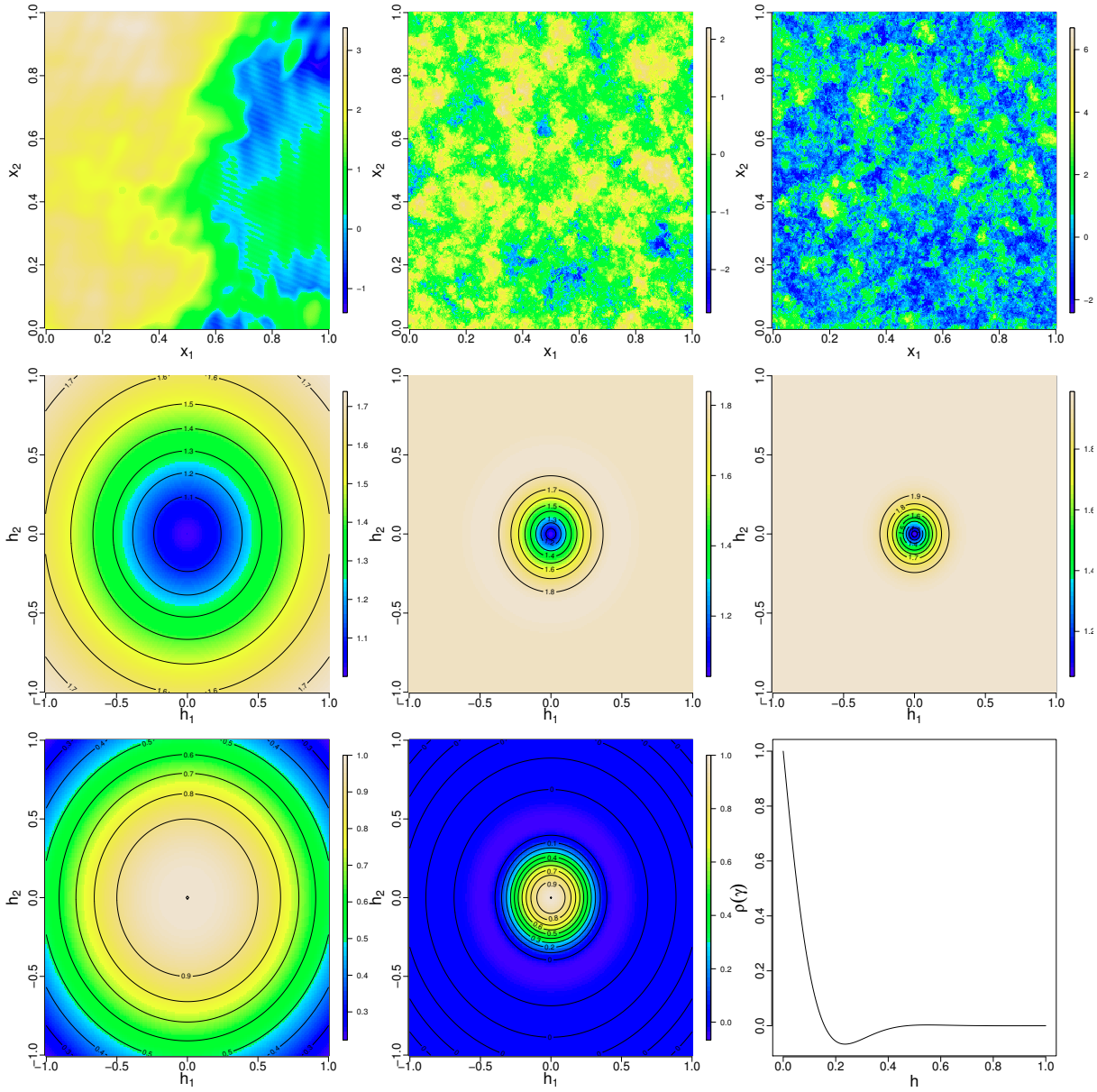

Figure 2.4: Representation of Extremal-t processes on $\mathbb{S} = [0,1]^2$. The top row represents some realizations of the process while the second and third row shows respectively the extremal coefficient and the correlation function. The first column considers a Whittle-Matérn correlation function $\rho(h)$ with smoothness parameter 0.8 and degree of freedom $\nu = 2$. The second and third column display the process with degrees of freedom respectively $\nu = 2$ and 8 and the same damped cosine correlation function $\rho(h)$ with scale $\lambda = 0.1$ and smoothness 1. The middle and right panels of the bottom row respectively provide the two and one-dimensional representations of the isotropic correlation function. The margins are transformed to be Gumbel.

class of max-stable process which contains most of the models listed above as special cases. Its spectral construction is provided and some of its properties are studied.

Chapter 3

On the bivariate Brown-Resnick max-stable processes with spatial dependence parameter

3.1 Motivation

Over the last decade, max-stable processes have been intensively applied to spatial environmental events; see for example Buishand et al. (2008); Padoan et al. (2010); Blanchet and Davison (2011); Engelke et al. (2012), Genton et al. (2015). However, in practice, such spatial events often depend on other variables. Genton et al. (2015) give an extension of max-stable processes to the multivariate setting.

The Brown-Resnick process (Brown and Resnick (1977) and Kabluchko et al. (2009)) discussed in Section 2.6.3 has received a particular interest in this context. The first authors to work with a bivariate version of the Brown-Resnick process are de Haan and Zhou (2008) and Buishand et al. (2008). The spatial components of the model are split such that they can be considered as separate variables. An application to rainfall over the region of North Holland in the Netherlands is given by Buishand et al. (2008). The aim of this analysis is to give an overview of the global rainfall over the study region which can be taken into account to calculate return levels.

A multivariate representation of Brown-Resnick processes is given by Molchanov and Stucki (2013) and Oesting et al. (2013) give an application, in the bivariate context, to wind gusts using the association between observable data and the corresponding forecast.

In this chapter the bivariate process defined by de Haan and Zhou (2008) and Buishand et al. (2008) is taken into account. The balance between the simplicity of the model, its easy simulation and its finesse are the main advantages of working with such process. Furthermore the shift stationarity of the process is also an advantageous property for most spatial environmental challenges.

The work presented in this chapter provides some extensions of the seminal works by Brown

and Resnick (1977); de Haan and Zhou (2008) and can also be considered in the generalised case of Kabluchko et al. (2009). A particular attention is drawn to the Brown-Resnick model introduced in Section 2.6.3. An additional general dependence parameter β is incorporated in the model (see de Haan and Pereira (2006)) to measure the quantity of spatial dependence. Unlike the indicators of spatial dependence introduced in Section 2.5 which consist in summarising the spatial dependence after fitting a max-stable model to the data, this parameter is considered as an intrinsic parameter of the model, the purpose being to model its effects on the simulation of extremes.

In order to use the process to analyse spatial extremes the spatial dependence parameter needs to be estimated. Ideally, estimation should be done by analyzing the sample paths of n processes but unfortunately, in practice, these are observed at a finite number of locations. In their application to rainfall in the North Holland region, Buishand et al. (2008) estimated the dependence parameter according to the method proposed by de Haan and Pereira (2006) which consists in averaging the $\binom{p}{2}$ pairwise dependence estimators between all p recording stations.

The spatial dependence parameter can be given as a function of the pairwise tail dependence function for which a consistent and asymptotically normal non-parametric estimator is available (see Huang (1992); Drees and Huang (1998); Einmahl et al. (2008)). Einmahl et al. (2012) introduced the multivariate equivalent of the estimator of the tail dependence function. Consequently if the d-dimensional marginal distributions of the process are known, then the dependence parameter can be estimated straight away, without approximating it by the average of its pairwise components. An insightful comparison study would be to assess the quality of the parameter estimate when calculated by averaging pairwise coefficients or using the whole information at once. The efficiency comparison of the estimates should also take into account the computational load that both require. Additionally other estimates, based on triplewise, quadruplewise or n-tuplewise coefficients, should be considered.

Section 3.2 sets the framework and defines the univariate and bivariate max-stable processes as well as the construction of the unique parameter of the model. Results on the distribution of the processes are given in Section 3.3 and their exactness is proved by simulation in Section 3.4.

3.2 Background and definitions

Our starting point is the Brown-Resnick process based on the representation of Theorem 2.5 and which was detailed in Section 2.6.3. Denote by β the general dependence parameter, the univariate process is defined by

$$\{\eta'(s)\}_{s\in\mathbb{S}} := \left\{\max_{1\leq i} \zeta_i \exp\left(\epsilon_i(\beta s) - \frac{\beta|s|}{2}\right)\right\}_{s\in\mathbb{S}}, \tag{3.1}$$

where $\mathbb{S} \subset \mathbb{R}$. Appendix B proves the existence of this max-stable model and that the conditions of Theorem 2.5 are met. Furthermore the process is a simple max-stable (its margins are unit Fréchet) and, due to the stationarity of Ornstein-Ulhenbeck processes, it is also stationary. Small values of β refer to strong dependence whereas large values are associated to weak dependence.

In order to being able to explore the whole real line and not to be restricted to its positive part, the ϵ_i represent independent copies of a double-sided Brownian motion ϵ. Let B_1 and B_2 be independent Brownian motions then

$$\epsilon(s) := \begin{cases} B_1(s), & s \geq 0 \\ B_2(-s), & s < 0 \end{cases}.$$

de Haan and Zhou (2008) and Buishand et al. (2008) define a natural extension of this process to the two-dimensional context, for $s = (s_1, s_2) \in \mathbb{S} \subset \mathbb{R}^2$, by

$$\{\eta(s)\}_{s\in\mathbb{S}} = \left\{ \max_{i\geq 1} \zeta_i \exp\left(\epsilon_{1i}(\beta s_1) + \epsilon_{2i}(\beta s_2) - \frac{\beta}{2}(|s_1| + |s_2|) \right) \right\}_{s\in\mathbb{S}}, \tag{3.2}$$

where ϵ_{1i} and ϵ_{2i} are independent copies of the double-sided Brownian motion ϵ. For all $s_1, s_2 \in \mathbb{R}$, the bivariate process η given by (3.2) fulfills the conditions of Theorem 2.5:

$$\mathbb{E}\left[\exp\left\{ \epsilon_{1i}(\beta s_1) + \epsilon_{2i}(\beta s_2) - \frac{\beta}{2}(|s_1| + |s_2|) \right\} \right] = 1,$$

and for $a_1 < b_1$, $a_2 < b_2$, $a_1, a_2, b_1, b_2 \in \mathbb{R}$,

$$\mathbb{E}\left[\sup_{\substack{a_1\leq s_1\leq b_1 \\ a_2\leq s_2\leq b_2}} \exp\left\{ \epsilon_{1i}(\beta s_1) + \epsilon_{2i}(\beta s_2) - \frac{\beta}{2}(|s_1| + |s_2|) \right\} \right] < \infty.$$

The process η can be written as a combination of two one-dimensional processes similar to η'. Moreover, because the process η' can be obtained as the limit of the pointwise maxima of i.i.d. Ornstein-Ulhenbeck processes, η is stationary.

Definition 3.1 (de Haan and Zhou (2008))**.**

(i) The two-dimensional marginal distributions of η' are

$$-\log \Pr(\eta'(s_1) \leq e^x, \eta'(s_2) \leq e^y) = e^{-x}\Phi\left(\frac{\sqrt{\beta h'}}{2} + \frac{y-x}{\sqrt{\beta h'}} \right) + e^{-y}\Phi\left(\frac{\sqrt{\beta h'}}{2} + \frac{x-y}{\sqrt{\beta h'}} \right),$$

for $s_1, s_2 \in \mathbb{R}$, with $h' = |s_1 - s_2|$.

(ii) The two-dimensional marginal distributions of η are

$$-\log \Pr\left(\eta(s_1) \leq e^x, \eta(s_2) \leq e^y\right) = e^{-x}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{y-x}{\sqrt{\beta h}} \right) + e^{-y}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{x-y}{\sqrt{\beta h}} \right),$$

for $s_i = \left(s_i^{(1)}, s_i^{(2)}\right)^\top \in \mathbb{R}^2, i = 1, 2$, with $h = \left|s_1^{(1)} - s_2^{(1)}\right| + \left|s_1^{(2)} - s_2^{(2)}\right|$.

The distances h' and h represent the univariate and bivariate L_1 distance (also called Manhattan or Taxicab distance). The L_1 distance between two d-dimensional vectors s and t, in a

real vector space is given by

$$h := h(s,t) = \sum_{i=1}^{d} |s_i - t_i|.$$

According to these distributions, both models can be said to be shift stationary or invariant under a shift.

From the above definition, the amount of spatial dependence between two locations s_1 and s_2 in $\mathbb{R}^2$ can be evaluated by

$$\beta = \frac{4}{h}\left(\Phi^{\leftarrow}\left(-\frac{1}{2}\log \Pr(\eta(s_1) \leq 1, \eta(s_2) \leq 1)\right)\right)^2,$$

where $\Phi^{\leftarrow}$ denotes the inverse of Φ. Note that this result holds also for the univariate process η' by substituting h by h'. The $-\log \Pr(\eta(s_1) \leq 1, \eta(s_2) \leq 1)$ term corresponds to the stable tail dependence function l at $(1,1)$ (see Chapter 1; Einmahl et al. (2008)) and thus can be estimated using the non-parametric estimator introduced by Huang (1992) (see also Drees and Huang (1998); Einmahl et al. (2008) and Einmahl et al. (2012)). Thus an estimate of β is obtained by plugging an estimator of l in (3.2).

Consider X and Y two random variables and denote by $\{X_{i,n}\}_{i=1}^n$ and $\{Y_{i,n}\}_{i=1}^n$ their order statistics. Huang (1992) and Drees and Huang (1998) propose the following estimator of the tail dependence function

$$\hat{l}_n(x,y) := \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \geq X_{n-kx+1,n} \text{ or } Y_i \geq Y_{n-ky+1,n}\}},$$

and Einmahl et al. (2008) introduce a similar estimator through

$$\hat{l}'_n(x,y) := \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \geq X_{n-kx+1/2,n} \text{ or } Y_i \geq Y_{n-ky+1/2,n},\}}$$

which shows slightly better estimators for simulations of finite samples and gives similar results for large samples. Both of these estimators are consistent under the condition that $k = k(n) \to \infty$, $k(n)/n \to 0$, $n \to \infty$ and are asymptotically normal if adding extra mild conditions.

3.3 Results

This section showcases our main results on the distributions of the univariate and bivariate Brown-Resnick process η and η'. Some preliminary results are listed before each of our contributions.

3.3.1 Preliminary results to Section 3.3.2

Azzalini (1985) introduced the Skew-Normal distribution which was extended to the multivariate case by Azzalini and Dalla Valle (1996). Later Capitanio et al. (2003) provided a broader

class of distributions which contains the Skew-Normal, called the extended Skew-Normal distribution and defined as follows

Definition 3.2. Consider the vectors $\epsilon = (\epsilon_1, \ldots, \epsilon_d)^\top \in \mathbb{R}^d$, $\epsilon^* = (\epsilon_0, \epsilon)^\top \in \mathbb{R}^{d+1}$ and let

$$\epsilon^* = (\epsilon_0, \epsilon)^\top \sim \mathcal{N}_{d+1}\left(0, \overline{\Omega}^*\right), \qquad \overline{\Omega}^* = \begin{pmatrix} 1 & \delta^\top \\ \delta & \overline{\Omega} \end{pmatrix}$$

where $\overline{\Omega}^*$ is a full rank correlation matrix.

Let $Z = (\epsilon | \epsilon_0 + \tau > 0)$ then its probability density function is

$$f_Z(z) = \phi_d\left(z; \overline{\Omega}\right) \Phi\left(\alpha_0 + \alpha^\top z\right) \Phi^{-1}(\tau)$$

where

$$\alpha_0 = \tau\left(1 - \delta^\top \overline{\Omega}^{-1} \delta\right)^{-1/2}, \qquad \alpha = \left(1 - \delta^\top \overline{\Omega}^{-1} \delta\right)^{-1/2} \overline{\Omega}^{-1} \delta.$$

Equivalently we have

$$\alpha_0 = \tau\left(1 + \alpha^\top \overline{\Omega}^{-1} \alpha\right)^{1/2}, \qquad \delta = \left(1 + \alpha^\top \overline{\Omega}^{-1} \alpha\right)^{-1/2} \overline{\Omega} \alpha.$$

The cumulative distribution function of Z is given by

$$F_Z(z) = \Phi_{d+1}\left(\left(\tau, z^\top\right)^\top; \tilde{\Omega}\right) \Phi^{-1}(\tau), \qquad \tilde{\Omega} = \begin{pmatrix} 1 & -\delta^\top \\ -\delta & \overline{\Omega} \end{pmatrix}.$$

The following lemma arises from the distribution of the extended Skew-Normal distribution given above.

Lemma 3.3. *Let $a = (a_1, \ldots, a_n)^\top$ and $b = (b_1, \ldots, b_n)^\top$ be two vectors of $\mathbb{R}^n$ and consider R_n a $n \times n$ correlation matrix where $[R]_{i,j} = \rho_{i,j}$ ($\rho_{i,j} = 1$ if $i = j$). Let $u \in \mathbb{R}$. Based on Definition 3.2 we have,*

$$\int_{-\infty}^{u} \phi(x) \Phi_k(a + bx; R_n)\, dx = \Phi_{k+1}\left(\left\{u, a(1 + b^2)^{-1/2}\right\}^\top; R_{n+1}\right),$$

where

$$R_{n+1} = \begin{bmatrix} 1 & -\left(\frac{b}{1+b^2}\right)^\top \\ -\left(\frac{b}{1+b^2}\right) & R'_n \end{bmatrix},$$

and where R'_n is a $n \times n$ correlation matrix with $[R'_n]_{i,j} = (b_i b_j + \rho_{i,j}) / \sqrt{(1 + b_i^2)(1 + b_j^2)}, i, j = 1, \ldots, k$.

Another somewhat important result is given in Abramowitz and Stegun (1964) deals with the decomposition of the bivariate normal pdf

Lemma 3.4. *Consider the zero mean bivariate normal pdf*

$$\phi_2(x,y;\rho) = (2\pi)^{-1}\left(1-\rho^2\right)^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right)\right\},$$

then it can be decomposed as

$$\phi_2(x,y;\rho) = \left(1-\rho^2\right)^{-1/2}\phi(x)\phi\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right).$$

3.3.2 Marginal distributions of univariate Brown-Resnick max-stable processes

The two results given in this section are about the marginal distribution of the univariate Brown-Resnick model η. First the 3-dimensional margins are detailed in order to show that it is possible to extend the result given in Definition 3.1 and secondly a mathematical induction is carried to establish the form of the d-dimensional margins.

Let $a_{i,j} = (\beta|s_j - s_i|)^{1/2}$ and denote by $\Phi_d(x;\Sigma)$ the d-dimensional centered normal cdf with variance-covariance matrix Σ of size $d \times d$. If Σ is a correlation matrix then a vector of correlation coefficients of size $\binom{n}{2}$ will be provided.

Theorem 3.5. *Suppose* $\{\eta(s)\}_{s\in\mathbb{S}}$ *is defined as in* (3.1). *Then for* $x_1, x_2, x_3 \in \mathbb{R}$ *and any* $s_1, s_2, s_3 \in \mathbb{R}$, *the three-dimensional distribution function of* $(\eta(s_1), \eta(s_2), \eta(s_3))$ *is given through*

$$\begin{aligned}
&-\log Pr(\eta(s_1) \le e^{x_1}, \eta(s_2) \le e^{x_2}, \eta(s_3) \le e^{x_3}) \\
&\qquad = e^{-x_1}\Phi_2\left(\left\{(x_3-x_1)a_{3,1}^{-1} + \tfrac{1}{2}a_{3,1}, (x_2-x_1)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right\}^\top; \rho_{2,3}^1\right) \\
&\qquad\quad + e^{-x_2}\Phi_2\left(\left\{(x_3-x_2)a_{3,2}^{-1} + \tfrac{1}{2}a_{3,2}, (x_1-x_2)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right\}^\top; \rho_{1,3}^2\right) \\
&\qquad\quad + e^{-x_3}\Phi_2\left(\left\{(x_2-x_3)a_{3,2}^{-1} + \tfrac{1}{2}a_{3,2}, (x_1-x_3)a_{3,1}^{-1} + \tfrac{1}{2}a_{3,1}\right\}^\top; \rho_{1,2}^3\right)
\end{aligned}$$

where, for $i, j, k = 1, 2, 3$, *the correlation coefficients* $\rho_{j,k}^i$ *are*

$$\rho_{j,k}^{(i)} = \begin{cases} 0 & \text{if } s_j < s_i < s_k \text{ or } s_j > s_i > s_k; \\ 1 & \text{if } s_j = s_k \\ \frac{\min(a_{i,j}, a_{i,k})^2}{a_{i,j}a_{i,k}} & \text{otherwise.} \end{cases}$$

The proof of Theorem 3.5 involves lengthy but easy calculations which will be detailed in Appendix C.1.

Theorem 3.5 proves that an analytical form of the marginal distributions of η' can be derived through some tedious but somewhat simple calculations. Two induction arguments are used to show that the marginal distributions can actually be given for any dimensions.

Lemma 3.6. *Suppose* $x = (x_1, \ldots, x_p)^\top$, $s = (s_1, \ldots, s_p)^\top \in \mathbb{R}^p$ *such that* $s_1 \le \ldots \le s_p$ *and let* $\epsilon(s_i)$ *denote a Brownian motion at time* s_i. *Suppose that* $Z = \exp\{\epsilon_{n,1} - \frac{1}{2}s_{n,1} - x_n\}$ *occurs*

with probability

$$Pr(\epsilon_{n,1} \geq x_{n,1} + \tfrac{1}{2}s_{n,1}, \ldots, \epsilon_{n,n-1} \geq x_{n,n-1} + \tfrac{1}{2}s_{n,n-1}, \\ \epsilon_{n+1,n} \leq x_{n+1,n} + \tfrac{1}{2}s_{n+1,n}, \ldots, \epsilon_{p,n} \leq x_{p,n} + \tfrac{1}{2}s_{p,n})$$

for $2 \leq n \leq p$ *and where* $\epsilon_{i,j} = \epsilon(s_i) - \epsilon(s_j)$, $x_{i,j} = x_i - x_j$, $s_{i,j} = s_i - s_j$, *then*

$$\begin{aligned}\mathbb{E}[Z] &= e^{-x_n} Pr(\epsilon_{n,1} \leq x_{1,n} + \tfrac{1}{2}s_{n,1}, \ldots, \epsilon_{n,n-1} \leq x_{n-1,n} + \tfrac{1}{2}s_{n,n-1}, \\ &\qquad \epsilon_{n+1,n} \leq x_{n+1,n} + \tfrac{1}{2}s_{n+1,n}, \ldots, \epsilon_{p,n} \leq x_{p,n} + \tfrac{1}{2}s_{p,n}) \\ &= e^{-x_n}\Phi_{p-1}\left((x_{-n} - x_n)|s_{-n} - s_n|^{-1/2} + |s_{-n} - s_n|^{1/2}/2; R\right),\end{aligned}$$

where R *is a correlation matrix,* $x_{-i} = x\backslash\{x_i\}$ *and* $s_{-i} = s\backslash\{s_i\}$ *meaning that the* i*-th elements of* x *and* s *are discarded.*

Refer to Appendix C.2 for a detailed proof and the analytical form of R.

Theorem 3.7. *Suppose* $x = (x_1, \ldots, x_n)^\top, s = (s_1, \ldots, s_n)^\top \in \mathbb{R}^n$. *Denote* $x_{-i} = x\backslash\{x_i\}$, $s_{-i} = s\backslash\{s_i\}$ *and* $a_{-i,j} = (a_{1,j}, \ldots, a_{i-1,j}, a_{i+1,j}, \ldots, a_{n,j})^\top$ *where* $a_{i,j} = (\beta|s_j - s_i|)^{1/2}$. *Then the* n*-dimensional marginal distribution of* $\{\eta'(s)\}_{s\in\mathcal{S}}$ *is given by*

$$-\log Pr(\eta'(s_1) \leq e^{x_1}, \ldots, \eta(s_n) \leq e^{x_n}) = \sum_{j=1}^{n} e^{-x_i}\Phi_{n-1}\left(q^{(j)}; R^{(j)}\right),$$

where $q^{(j)} = (x_{-j} - x_j)a_{-j,j}^{-1} + a_{-j,j}/2$ *and the correlation matrix is given by* $R^{(j)} = \left[R^{(j)}_{k,l}\right]_{k,l=\{1,\ldots,n\}\backslash\{j\}}$ *with*

$$R^{(j)}_{k,l} = \begin{cases} 0 & \text{if } \min(s_k, s_l) < s_j < \max(s_k, s_l), \\ 1 & \text{if } s_k = s_l, \\ \frac{\min(a_{j,k}, a_{j,l})}{\max(a_{j,k}, a_{j,l})} = \frac{\min(a_{j,k}, a_{j,l})^2}{a_{j,k}a_{j,l}} & \text{otherwise.} \end{cases}$$

See Appendix C.3 for a proof.

We have demonstrated that it is possible to give the analytical form of the marginal distributions of the univariate Brown-Resnick process. The second part of this section shows that it is possible to provide similar results for the bivariate process. Again some preliminary results are required.

3.3.3 Preliminary results to Section 3.3.4

Lemma 3.8. *Suppose* ϵ *and* ϵ' *are normally distributed random variables with zero mean and variances respectively* u *and* v. *For* $a, c \neq 0, b$ *and* d *constants, then*

(i) If $\epsilon = \epsilon' + \epsilon''$ *where* ϵ'' *is independent of* ϵ' *with mean* 0 *and variance* $u - v$ *then*

$$\mathbb{E}e^{\epsilon - u/2}\Phi_2\left(\{a\epsilon + b, c\epsilon' + d\}^\top; \rho\right) = \Phi_2\left(\left\{\frac{au + b}{\sqrt{1 + a^2u}}, \frac{cv + d}{\sqrt{1 + c^2v}}\right\}^\top; \rho^*\right),$$

where $\rho^* = (\rho + acv)((1+a^2u)(1+c^2v))^{-1/2}$.

(ii) If ϵ *and* ϵ' *are independent then*

$$\mathbb{E}e^{\epsilon - u/2}\Phi_2\left(\{a\epsilon + b, c\epsilon' + d\}^\top; \rho\right) = \Phi_2\left(\left\{\frac{au+b}{\sqrt{1+a^2u}}, \frac{d}{\sqrt{1+c^2v}}\right\}^\top; \rho^*\right),$$

where $\rho^* = \rho((1+a^2u)(1+c^2v))^{-1/2}$.

See Appendix C.4 for a proof.

Lemma 3.9. *Suppose* ϵ, ϵ' *and* ϵ'' *are normally distributed random variables with zero mean and variances respectively* u, v *and* w. *Suppose also that* ϵ *is independent of* ϵ' *and* ϵ'' *which are themselves correlated. For* $a, b, d, e \neq 0$ *and* c, f *constants, if* $\epsilon'' = \epsilon' + \epsilon'''$ *where* ϵ''' *is independent of* ϵ' *with mean* 0 *and variance* $w - v$, *then*

$$\mathbb{E}e^{\epsilon - u/2}\Phi_2\Big(\{a\epsilon + b\epsilon' + c, d\epsilon + e\epsilon'' + f\}^\top; \rho\Big) = \Phi_2\left(\left\{\frac{au+c}{\sqrt{1+a^2u+b^2v}}, \frac{du+f}{\sqrt{1+d^2u+e^2w}}\right\}^\top; \rho^*\right),$$

where $\rho^* = (\rho + adu + bev)((1+a^2u+b^2v)(1+d^2u+e^2w))^{-1/2}$.

See Appendix C.5 for a proof.

Lemma 3.10. *Suppose* ϵ, ϵ' *and* ϵ'' *are normally distributed and mutually independent random variables with zero mean and variances respectively* u, v *and* w. *For* $a, b, c, e \neq 0$ *and* d, f *constants, then*

$$\mathbb{E}\, e^{\epsilon + \epsilon' - (u+v)/2}\Phi_2\Big(\{a\epsilon + b\epsilon' + c\epsilon'' + d, e\epsilon + f\}^\top; \rho\Big) = \Phi_2\left(\left\{\frac{au+bv+d}{\sqrt{1+a^2u+b^2v+c^2w}}, \frac{eu+f}{\sqrt{1+e^2u}}\right\}^\top; \rho^*\right),$$

where $\rho^* = (\rho + aeu)((1+a^2u+b^2v+c^2w)(1+e^2u))^{-1/2}$.

See Appendix C.6 for a proof.

3.3.4 Marginal distributions of bivariate Brown-Resnick max-stable processes

In this section we will consider the process $\{\eta(u)\}$, $s = \left(s^{(1)}, s^{(2)}\right)^\top \in \mathbb{R}^2$, introduced in (3.2) and define the expression of its multi-dimensional distribution functions. Throughout this section we will make use of the following notations

$$\begin{aligned}
&s_i = \left(s_i^{(1)}, s_i^{(2)}\right)^\top \in \mathbb{R}^2\\
&h_{i,j}^{(k)} = \beta|s_i^{(k)} - s_j^{(k)}|, \quad k = 1, 2\\
&h_{i,j} = h_{i,j}^{(1)} + h_{i,j}^{(2)} = \beta|s_i^{(1)} - s_j^{(1)}| + \beta|s_i^{(2)} - s_j^{(2)}|\\
&g_{i,j} \equiv g(s_i, s_j) = (h_{i,j})^{1/2}.
\end{aligned}$$

The following result gives the three-dimensional marginal distribution function of the bivariate process η' and helps us to understand the change in the correlation matrix when using bivariate processes instead of univariate processes.

Theorem 3.11. *Suppose $x_1, x_2, x_3 \in \mathbb{R}$ and $s_1, s_2, s_3 \in \mathbb{R}^2$, the three-dimensional distribution function of $(\eta(s_1), \eta(s_2), \eta(s_3))$ is given by*

$$\begin{aligned}
&-\log Pr(\eta(s_1) \leq e^{x_1}, \eta(s_2) \leq e^{x_2}, \eta(s_3) \leq e^{x_3}) \\
&\qquad = e^{-x_1}\Phi_2\left(\left\{(x_2-x_1)g_{1,2}^{-1} + \tfrac{1}{2}g_{1,2}, (x_3-x_1)g_{1,3}^{-1} + \tfrac{1}{2}g_{1,3}\right\}^\top; \rho_{2,3}^1\right) \\
&\qquad\quad + e^{-x_2}\Phi_2\left(\left\{(x_1-x_2)g_{1,2}^{-1} + \tfrac{1}{2}g_{1,2}, (x_3-x_2)g_{2,3}^{-1} + \tfrac{1}{2}g_{2,3}\right\}^\top; \rho_{1,3}^2\right) \\
&\qquad\quad + e^{-x_3}\Phi_2\left(\left\{(x_1-x_3)g_{1,3}^{-1} + \tfrac{1}{2}g_{1,3}, (x_2-x_3)g_{2,3}^{-1} + \tfrac{1}{2}g_{2,3}\right\}^\top; \rho_{1,2}^3\right),
\end{aligned}$$

where, for $i, j, k = 1, 2, 3$, the correlation coefficients $\rho_{j,k}^i$ are

$$\rho_{j,k}^{(i)} = \begin{cases} 0 & \text{if } \min\left(s_j^{(l)}, s_k^{(l)}\right) < s_i^{(l)} < \max\left(s_j^{(l)}, s_k^{(l)}\right), \forall l = 1, 2 \\ 1 & \text{if } s_j = s_k \\ \frac{\min\left(\beta h_{i,j}^{(1)}, \beta h_{i,k}^{(1)}\right) + \min\left(\beta h_{i,j}^{(2)}, \beta h_{i,k}^{(2)}\right)}{g_{i,j} g_{i,k}} & \text{otherwise.} \end{cases}$$

All the tools required to prove this theorem are available above and details of the calculations can be found in Appendix C.7. The comparison of the distributions of the processes η' and η given in Theorem 3.5 and Theorem 3.11, draws the same conclusions as the ones already made when comparing the results of de Haan and Zhou (2008) and summarised in Definition 3.1: both are identical functions of the distance h between two locations and they only differ by the expression of the L_1 distance.

We are now able to extrapolate the n-dimensional marginal distribution of bivariate Brown-Resnick max-stable processes, through the following theorem:

Theorem 3.12. *Suppose $x = (x_1, \ldots, x_n)^\top, s = (s_1, \ldots, s_n)^\top \in \mathbb{R}^n$. Denote $x_{-i} = x \backslash \{x_i\}$, $s_{-i} = s \backslash \{s_i\}$ and $g_{-i,j} = (g_{1,j}, \ldots, g_{i-1,j}, g_{i+1,j}, \ldots, g_{n,j})^\top$. Then the n-dimensional marginal distribution of $\{\eta(s)\}_{s \in \mathcal{S}}$ is given by*

$$-\log Pr(\eta(s_1) \leq e^{x_1}, \ldots, \eta(s_n) \leq e^{x_n}) = \sum_{j=1}^{n} e^{-x_i}\Phi_{n-1}\left(q^{(j)}; R^{(j)}\right)$$

where $q^{(j)} = (x_{-j} - x_j)g_{-j,j}^{-1} + g_{-j,j}/2$ and the correlation matrix is given by $R^{(j)} = \left[R_{k,l}^{(j)}\right]_{k,l=\{1,\ldots,n\}\backslash\{j\}}$ with

$$R_{j,k}^{(i)} = \begin{cases} 0 & \text{if } \min\left(s_k^{(m)}, s_l^{(m)}\right) < s_j^{(m)} < \max\left(s_k^{(m)}, s_l^{(m)}\right), \forall m = 1, 2 \\ 1 & \text{if } s_k = s_l \\ \frac{\min\left(h_{j,k}^{(1)}, h_{j,l}^{(1)}\right) + \min\left(h_{j,k}^{(2)}, h_{j,l}^{(2)}\right)}{g_{j,k} g_{j,l}} & \text{otherwise.} \end{cases}$$

To prove Theorem 3.12, it is natural to use the proof of the n-dimensional result for univariate processes given in Theorem 3.7 where the distance between two points $a_{i,j}$ is replaced by the distance $g_{i,j}$. The proof of Theorem 3.12 relies on a very similar induction argument to its univariate equivalent, which is omitted here.

An important difference between the univariate and bivariate case is the expression of the correlation coefficient $R^{(i)}_{j,k}$. If the denominator follows the same change of metric, the numerator is now considering the sum of componentwise minima. Both the minima in the first and second components, for all the locations involved, are involved in the calculation of each correlation coefficient. Moreover it is clear that if either the first or the second component are not considered Theorem 3.7 would be recovered.

The next section provide two simulation studies in order to support the exactness of Theorem 3.7 and Theorem 3.12.

3.4 Simulation results for high-dimensional marginal distribution of Brown-Resnick max-stable processes

In this section we provide simulation results for the marginal distributions of univariate and bivariate Brown-Resnick process, in $2, 3, 4, 5, 6$ and 7 dimensions.

Consider the two Brown-Resnick processes defined respectively in (3.1) and (3.2) with $\zeta_i = 1/(E_1 + E_2 + \cdots + E_i)$ where E_i's are i.i.d. standard exponential random variables.

Locations in space are randomly selected (in both $\mathbb{R}$ and $\mathbb{R}^2$). For simplicity the spatial dependence parameter β is set to 1, and then each processes are simulated 1000 times at $2, 3, 4, 5, 6$ and 7 of these locations. We suppose $z = e^{-x_i}, i = 1, \cdots, 7$ in order to evaluate the empirical and theoretical estimates of the marginal distributions. In other words, the joint probability that the processes are less than or equal to the same critical value z. A total of 100 values of z, equally spaced between 0.1 and 10 are considered.

Results are given in Figure 3.1 for the univariate process and in Figure 3.2 for the bivariate process.

The empirical estimate of the marginal distribution consist in calculating the average of times (out of 1000) that all the processes are taking a value less than or equal to z. In Figure 3.1 and Figure 3.2, the grey shaded area represents the 95% pointwise confidence band which are calculated from some value z by

$$\left[\text{mean}_e(z) \pm \frac{\text{mean}_e(z)(1 - \text{mean}_e(z))}{n - 1}\right]$$

where $\text{mean}_e(z)$ denotes the empirical mean at z and n is the sample size ($n = 1000$). True values are represented by solid black lines and calculated using the results of Theorem 3.7 and Theorem 3.12.

In Figure 3.1, the true marginal distribution of univariate processes does lie within the shaded area for all the dimensions considered. Processes were simulated using the `R` package

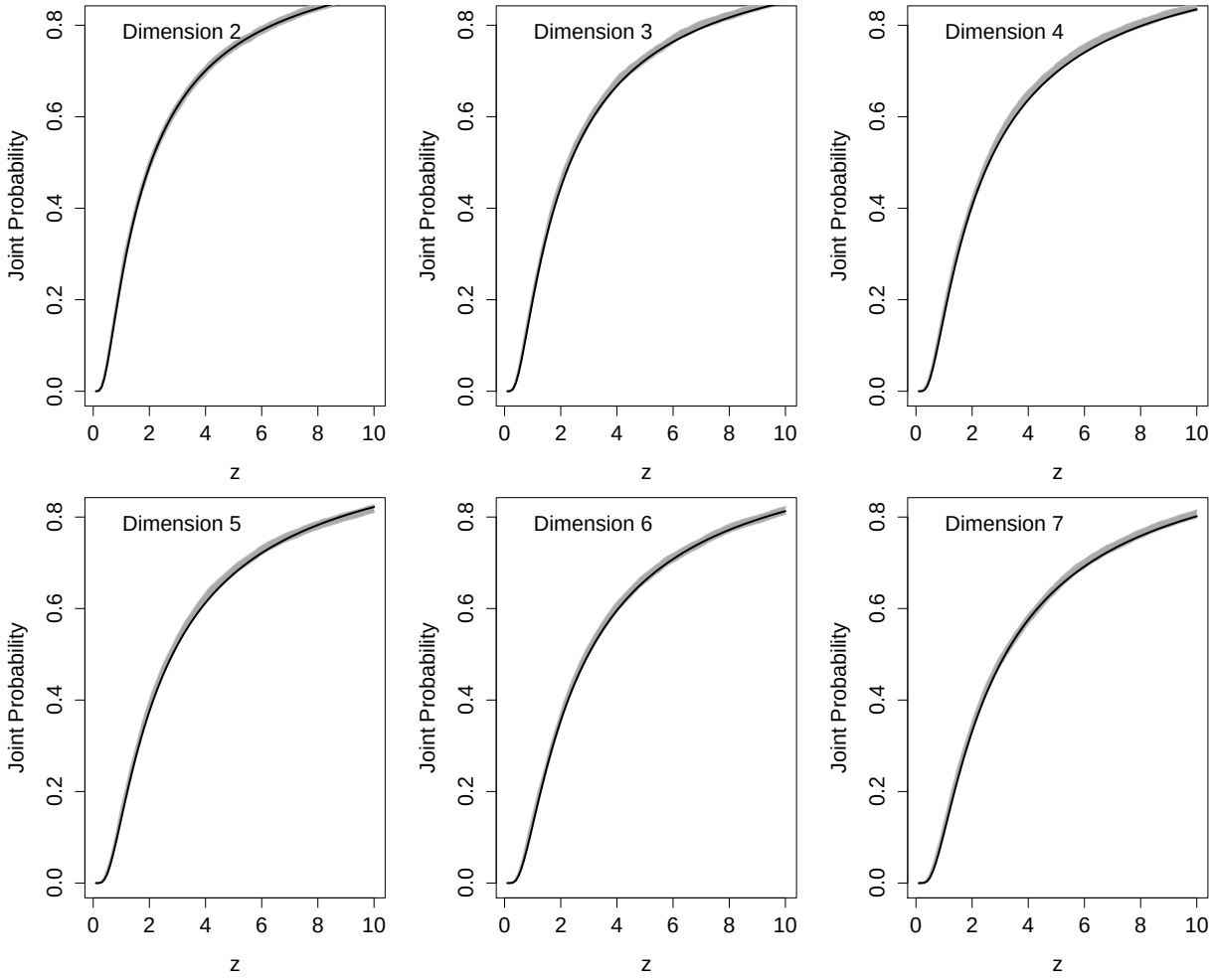

Figure 3.1: 2, 3, 4, 5, 6 and 7-dimensional marginal distributions of the univariate Brown-Resnick max-stable process. The grey shaded area corresponds to the 95% pointwise confidence band for the empirical estimates. The black solid line provides the theoretical value.

`SpatialExtremes` (Ribatet et al. (2013)). The width of the confidence bands is relatively small meaning that the simulation procedure is very robust, leading only to small variations. Moreover the true joint probability is always within the band, meaning that the theoretical result is verified.

In Figure 3.2, the true joint probability is contained within the 95% confidence interval for low dimensions. However, as the dimension increases, the distribution seems to be slightly over estimated for small value of z. This may be due to the quality of the simulation procedure which isn't as straight forward as in the univariate case.

3.5 Conclusion and remarks

New asymptotic results on bivariate Brown-Resnick processes have been developed. The univariate process has been extended to the bivariate case and a general dependence structure is taken into account (see de Haan and Pereira (2006); de Haan and Zhou (2008); Buishand

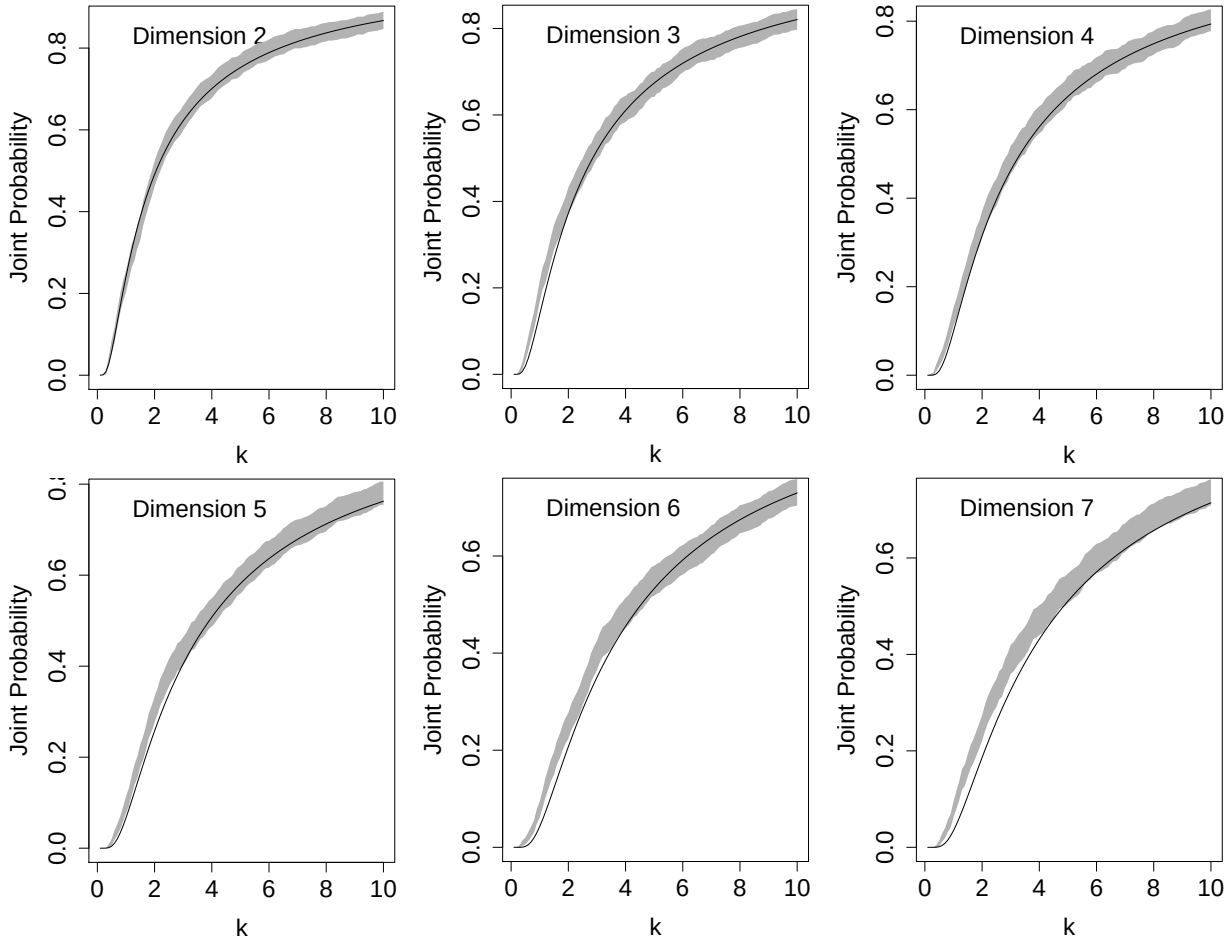

Figure 3.2: 2, 3, 4, 5, 6 and 7-dimensional marginal distributions of the bivariate Brown-Resnick max-stable process. The grey shaded area corresponds to the 95% pointwise confidence band for the empirical estimates. The black curve gives the theoretical value.

et al. (2008)). We derived the high-dimensional marginal distributions for both processes, allowing easy estimation of the spatial dependence. Using empirical estimators from Einmahl et al. (2012) the joint distribution can be estimated to derive an estimate of the spatial dependence. Buishand et al. (2008) considered the average of $\binom{p}{2}$ pairwise dependences (i.e. using the pairwise marginal distribution function) and this method should have been compared to other estimation procedures such as directly considering the p-dimensional distributions or other high order distributions.

While these results were developed, we have become aware that Huser and Davison (2013) derived at the same time the multivariate Brown-Resnick distribution in the univariate case. At the moment our results are not ready for publication and extra work is needed to provide some novelty to the present literature. In their work Huser and Davison (2013) provide a proof for high order margins of the generalized Brown-Resnick process (see (2.20)). Clearly the results from Theorem 3.7 and Theorem 3.12 can be recovered from their results by taking the right semi-variogram. They considered pairwise likelihood approaches (pairwise and triplewise) to fit

the max-stable process. It was found that there is no gain in efficiency to consider higher order likelihood hence that pairwise composite likelihood was the most appropriate method to use (it is also less computationally demanding). To conclude we would expect then similar results in the setup of this Chapter, meaning that Buishand et al. (2008) used an efficient procedure to estimate the spatial dependence despite being limited by only knowing the 2-dimensional result from de Haan and Zhou (2008). Finally, a pattern has appeared when going from univariate processes to bivariate processes, it seems then straightforward to establish the n-dimensional marginal distribution of d-dimensional processes. It could allow to consider multiple variables and to investigate about their dependence as suggested by Genton et al. (2015) and Oesting et al. (2013).

Chapter 4

Models for extremal dependence derived from skew-symmetric families

Abstract

Skew-symmetric families of distributions such as the skew-normal and skew-t represent supersets of the normal and t distributions, and they exhibit richer classes of extremal behaviour. By defining a non-stationary skew-normal process, which allows the easy handling of positive definite, non-stationary covariance functions, we derive a new family of max-stable processes – the extremal-skew-t process. This process is a superset of non-stationary processes that include the stationary extremal-t processes. We provide the spectral representation and the resulting angular densities of the extremal-skew-t process, and illustrate its practical implementation

Keywords: Asymptotic independence; Angular density; Extremal coefficient; Extreme values; Max-stable distribution; Non-central extended skew-t distribution; Non-stationarity; Skew-Normal distribution; Skew-Normal process; Skew-t distribution.

4.1 Introduction

The modern-day analysis of extremes is based on results from the theory of stochastic processes. In particular, max-stable processes (de Haan, 1984) are a popular and useful tool when modelling extremal responses in environmental, financial and engineering applications. Let $\mathbb{S} \subseteq \mathbb{R}^k$ denote a k-dimensional region of space (or space-time) over which a real-valued stochastic process $\{Y(s)\}_{s\in\mathbb{S}}$ with a continuous sample path on $\mathbb{S}$ can be defined. Considering a sequence $Y_1, \ldots, Y_n$ of i.i.d. copies of Y, the pointwise partial maximum can be defined as

$$M_n(s) = \max_{i=1,\ldots,n} Y_i(s), \quad s \in \mathbb{S}.$$

If there are sequences of real-valued functions, $a_n(s) > 0$ and $b_n(s)$, for $s \in \mathbb{S}$ and $n = 1, 2, \ldots$, such that

$$\left\{ \frac{M_n(s) - b_n(s)}{a_n(s)} \right\}_{s \in \mathbb{S}} \Rightarrow \{U(s)\}_{s \in \mathbb{S}},$$

converges weakly as $n \to \infty$ to a process $U(s)$ with non-degenerate marginal distributions for all $s \in \mathbb{S}$, then $U(s)$ is known as a max-stable process (de Haan and Ferreira, 2006, Ch. 9). In this setting, for a finite sequence of points $(s_j)_{j \in I}$ in some index set $I = \{1, \ldots, d\}$, the finite-dimensional distribution of U is then a multivariate extreme value distribution (de Haan and Ferreira, 2006, Ch. 6). This distribution has generalised extreme value univariate margins and, when parameterised with unit Fréchet margins, has a joint distribution function of the form

$$G(x_j, j \in I) = \exp\{-V(x_j, j \in I)\}, \quad x_j > 0,$$

where $x_j \equiv x(s_j)$. The exponent function V describes the dependence between extremes, and can be expressed as

$$V(x_j, j \in I) = \int_{\mathbb{W}} \max_{j \in I}(w_j / x_j) H(\mathrm{d}w_1, \ldots, \mathrm{d}w_d),$$

where the angular measure H is a finite measure defined on the d-dimensional unit simplex $\mathbb{W} = \{w \in \mathbb{R}^d : w_1 + \cdots + w_d = 1\}$, satisfying the moment conditions $\int_{\mathbb{W}} w_j \, H(\mathrm{d}w) = 1, j \in I$, (de Haan and Ferreira, 2006, Ch. 6).

In recent years a variety of specific max-stable processes have been developed, many of which have become popular as they can be practically amenable to statistical modelling (Davison et al., 2012). The extremal-t process (Opitz, 2013) is one of the best-known and widely-used max-stable processes, from which the Brown-Resnick process (Brown and Resnick, 1977; Kabluchko et al., 2009), the Gaussian extreme-value process (Smith, 1990) and the extremal-Gaussian processes (Schlather, 2002) can be seen as special cases. In their most basic form, the Brown-Resnick and the extremal-t processes can be respectively understood as the limiting extremal processes of strictly stationary Gaussian and Student-t processes. However, in practice, data may be non-stationary and exhibit asymmetric distributions in many applications. In these scenarios, skew-symmetric distributions (Azzalini, 2013; Arellano-Valle and Azzalini, 2006; Azzalini, 2005; Genton, 2004; Azzalini, 1985) provide simple models for modelling asymmetrically distributed data. However, the limiting extremal behaviour of these processes has not yet been established.

In this paper we characterise and develop statistical models for the extremal behaviour of skew-normal and skew-t distributions. The joint tail behaviours of these skew distributions are capable of describing a far wider range of dependence levels than that obtained under the symmetric normal and t distributions. We provide a definition of a skew-normal process which is in turn a non-stationary process. This provides an accessible approach to constructing positive definite, non-stationary covariance functions when working with non-Gaussian processes. Recently some forms of non-stationary dependent structures embedded into max-stable processes have been studied by Huser and Genton (2015). We show that on the basis of the skew-normal

process a new family of max-stable processes – the extremal-skew-t process – can be obtained. This process is a superset of non-stationary processes that includes the stationary extremal-t processes (Opitz, 2013). From the extremal-skew-t process, a rich family of non-stationary, isotropic or anisotropic extremal coefficient functions can be obtained.

This paper is organised as follows: in Section 4.2 we first introduce a new variant of the extended skew-t class of distributions, before developing a non-stationary version of the skew-normal process. In both cases we discuss the stochastic behavior of their extreme values. In Section 4.3 we derive the spectral representation of the extended extremal skew-t process. Section 4.4 discusses inferential aspects of the extremal skew-t dependence model, and Section 4.5 provides a real data application. We conclude with a Discussion.

4.2 Preliminary results on skew-normal processes and skew-t distributions

We introduce two preliminary results that will be used in order to present our main contribution in Section 4.3, the extremal-skew-t process. In Section 4.2.1 we define the *non-central* extended skew-t family of distributions, which is a new variant of the class introduced by Arellano-Valle and Genton (2010), that allows a non-central parameter. In Section 4.2.2 we present the development of a new non-stationary, skew normal random process.

Hereafter, we use $Y \sim \mathcal{D}_d(\theta_1, \theta_2, \ldots)$ to denote that Y is a d-dimensional random vector with probability law $\mathcal{D}$ and parameters $\theta_1, \theta_2, \ldots$. When $d = 1$ the subscript is omitted for brevity. Similarly, when a parameter is equal to zero or a scale matrix is equal to the identity (both in a vector and scalar sense) so that $\mathcal{D}_d$ reduces to an obvious sub-family, it is also omitted.

4.2.1 The non-central, extended skew-t distribution

While several skew-symmetric distributions have been developed (see e.g. Genton (2004) and Azzalini (2013)), we focus on the skew-normal and skew-t distributions.

Denote a d-dimensional skew-normally distributed random vector by $Y \sim \mathcal{SN}_d(\mu, \Omega, \alpha, \tau)$ (Arellano-Valle and Genton, 2010). This random vector has probability density function (pdf)

$$\phi_d(y; \mu, \Omega, \alpha, \tau) = \frac{\phi_d(y; \mu, \Omega)}{\Phi\{\tau/\sqrt{1 + Q_{\bar{\Omega}}(\alpha)}\}} \Phi(\alpha^\top z + \tau), \qquad y \in \mathbb{R}^d, \tag{4.1}$$

where $\phi_d(y; \mu, \Omega)$ is a d-dimensional normal pdf with mean $\mu \in \mathbb{R}^d$ and $d \times d$ covariance matrix Ω, $z = (y - \mu)/\omega$, $\omega = \mathrm{diag}(\Omega)^{1/2}$, $\bar{\Omega} = \omega^{-1}\,\Omega\,\omega^{-1}$, $Q_{\bar{\Omega}}(\alpha) = \alpha^\top \bar{\Omega} \alpha$ and $\Phi(\cdot)$ is the standard univariate normal cdf. The shape parameters $\alpha \in \mathbb{R}^d$ and $\tau \in \mathbb{R}$ are respectively *slant* and *extension* parameters. The cdf associated with (4.1) is termed the extended skew-normal distribution (Arellano-Valle and Genton, 2010) of which the skew-normal and normal distributions are special cases (Arellano-Valle and Genton, 2010; Azzalini, 2013). For example, in the case where $\alpha = 0$ and $\tau = 0$ the standard normal pdf is recovered.

Definition 4.1. Y is a d-dimensional, non-central extended skew-t distributed random vector, denoted by $Y \sim \mathcal{ST}_d(\mu, \Omega, \alpha, \tau, \kappa, \nu)$, if for $y \in \mathbb{R}^d$ it has pdf

$$\psi_d(y; \mu, \Omega, \alpha, \tau, \kappa, \nu) = \frac{\psi_d(y; \mu, \Omega, \nu)}{\Psi\left(\frac{\tau}{\sqrt{1+Q_{\bar{\Omega}}(\alpha)}}; \frac{\kappa}{\sqrt{1+Q_{\bar{\Omega}}(\alpha)}}, \nu\right)} \Psi\left\{(\alpha^\top z + \tau)\sqrt{\frac{\nu+d}{\nu + Q_{\bar{\Omega}^{-1}}(z)}}; \kappa, \nu+d\right\}, \tag{4.2}$$

where $\psi_d(y; \mu, \Omega, \nu)$ is the pdf of a d-dimensional t-distribution with location $\mu \in \mathbb{R}^d$, $d \times d$ scale matrix Ω and $\nu \in \mathbb{R}^+$ degrees of freedom, $\Psi(\cdot; a, \nu)$ denotes a univariate non-central t cdf with non-centrality parameter $a \in \mathbb{R}$ and ν degrees of freedom, and $Q_{\bar{\Omega}^{-1}}(z) = z^\top \bar{\Omega}^{-1} z$. The remaining terms are as defined in (4.1). The associated cdf is

$$\Psi_d(y; \mu, \Omega, \alpha, \tau, \kappa, \nu) = \frac{\Psi_{d+1}\{\bar{z}; \Omega^*, \kappa^*, \nu\}}{\Psi(\bar{\tau}; \bar{\kappa}, \nu)}, \tag{4.3}$$

where $\bar{z} = (z^\top, \bar{\tau})^\top$, Ψ_{d+1} is a $(d+1)$-dimensional (non-central) t cdf with covariance matrix and non-centrality parameters

$$\Omega^* = \begin{pmatrix} \bar{\Omega} & -\delta \\ -\delta^\top & 1 \end{pmatrix}, \quad \kappa^* = \begin{pmatrix} 0 \\ \bar{\kappa} \end{pmatrix},$$

and ν degrees of freedom, and where

$$\delta = \{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2} \bar{\Omega}\alpha, \quad \bar{\kappa} = \{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2} \kappa, \quad \bar{\tau} = \{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2} \tau. \tag{4.4}$$

When the non-centrality parameter κ is zero, then the extended skew-t family of Arellano-Valle and Genton (2010) is obtained. For the non-central skew-t family, we now demonstrate modified properties to those discussed in Arellano-Valle and Genton (2010).

Proposition 4.2 (Properties)**.** *Let $Y \sim \mathcal{ST}_d(\mu, \Omega, \alpha, \tau, \kappa, \nu)$.*

1. *Marginal and conditional distributions. Let $I \subset \{1, \ldots, d\}$ and $\bar{I} = \{1, \ldots, d\} \backslash I$ identify the d_I- and $d_{\bar{I}}$-dimensional subvector partition of Y such that $Y = (Y_I^\top, Y_{\bar{I}}^\top)^\top$, with corresponding partitions of the parameters (μ, Ω, α). Then*

   (a) *$Y_I \sim \mathcal{ST}_{d_I}(\mu_I, \Omega_{II}, \alpha_I^*, \tau_I^*, \kappa_I^*, \nu)$, where*

   $$\alpha_I^* = \frac{\alpha_I + \bar{\Omega}_{II}^{-1}\bar{\Omega}_{I\bar{I}}\alpha_{\bar{I}}}{\sqrt{1+Q_{\bar{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \quad \tau_I^* = \frac{\tau}{\sqrt{1+Q_{\bar{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \quad \kappa_I^* = \frac{\kappa}{\sqrt{1+Q_{\bar{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \tag{4.5}$$

   given $\tilde{\Omega}_{\bar{I}\bar{I}\cdot I} = \bar{\Omega}_{\bar{I}\bar{I}} - \bar{\Omega}_{\bar{I}I}\bar{\Omega}_{II}^{-1}\bar{\Omega}_{I\bar{I}}$.

   (b) *$(Y_{\bar{I}}|Y_I = y_I) \sim \mathcal{ST}_{d_{\bar{I}}}(\mu_{\bar{I}\cdot I}, \Omega_{\bar{I}\cdot I}, \alpha_{\bar{I}\cdot I}, \tau_{\bar{I}\cdot I}, \kappa_{\bar{I}\cdot I}, \nu_{\bar{I}\cdot I})$, where $\mu_{\bar{I}\cdot I} = \mu_{\bar{I}} + \Omega_{I\bar{I}}\Omega_{II}^{-1}(y_I - \mu_I)$, $\Omega_{\bar{I}\cdot I} = \zeta_I \Omega_{\bar{I}\bar{I}\cdot I}$, $\zeta_I = \{\nu + Q_{\Omega_{II}^{-1}}(z_I)\}/(\nu + d_I)$, $z_I = \omega_I^{-1}(y_I - \mu_I)$, $\omega_I = diag(\omega_{II})^{1/2}$, $Q_{\Omega_{II}^{-1}}(z_I) = z_I^\top \Omega_{II}^{-1} z_I$, $\Omega_{\bar{I}\bar{I}\cdot I} = \Omega_{\bar{I}\bar{I}} - \Omega_{\bar{I}I}\Omega_{II}^{-1}\Omega_{I\bar{I}}$, $\alpha_{\bar{I}\cdot I} = \omega_{\bar{I}\cdot I}\omega_{\bar{I}}^{-1}\alpha_{\bar{I}}$, $\omega_{\bar{I}\cdot I} = diag(\Omega_{\bar{I}\bar{I}\cdot I})^{1/2}$, $\omega_{\bar{I}} = diag(\omega_{\bar{I}\bar{I}})^{1/2}$, $\tau_{\bar{I}\cdot I} = \zeta_I^{-1/2}\{(\alpha_{\bar{I}}^\top \bar{\Omega}_{\bar{I}I}\bar{\Omega}_{II}^{-1} + \alpha_I^\top)z_I + \tau\}$, $\kappa_{\bar{I}\cdot I} = \zeta_I^{-1/2}\kappa$ and $\nu_{\bar{I}\cdot I} = \nu + d_I$.*

2. *Conditioning type stochastic representation. We can write* $Y = \mu + \Omega Z$, *where* $Z = (X|\alpha^\top X + \tau > X_0)$, *and where* $X \sim \mathcal{T}_d(\bar{\Omega}, \nu)$ *is independent of* $X_0 \sim \mathcal{T}(\kappa, \nu)$.

3. *Additive type stochastic representation. We can write* $Y = \mu + \Omega Z$, *where* $Z = \sqrt{\frac{\nu + \tilde{X}_0^2}{\nu+1}} X_1 + \delta \tilde{X}_0$, $X_1 \sim \mathcal{T}_d(\Omega - \delta\delta^\top, \bar{\kappa}, \nu + 1)$ *is independent of* $\tilde{X}_0 = (X_0 | X_0 + \bar{\tau} > 0)$, $X_0 \sim \mathcal{T}(\bar{\kappa}, \nu)$, $\delta \in (-1, 1)^d$ *and where* $\bar{\tau}$ *and* $\bar{\kappa}$ *are as in* (4.4).

Proof in Appendix D.1.1

We conclude by presenting a final property of the non-central skew-t family. The next result describes the extremal behaviour of observations drawn from a member of this class.

Proposition 4.3. *Let* $Z_1, \ldots, Z_n$ *be i.i.d. copies of* $Z \sim \mathcal{ST}_d(\bar{\Omega}, \alpha, \tau, \kappa, \nu)$ *and* M_n *be the componentwise sample maxima. Define* $a_n = (a_{n,1}, \ldots, a_{n,d})^\top$, *where*

$$a_{n,j} = \left\{ \frac{n\{\Gamma(\nu/2)\}^{-1}\Gamma\{(\nu+1)/2\}\nu^{(\nu-2)/2}\,\Psi(\alpha_j^*\sqrt{\nu+1}; \kappa, \nu+1)}{\sqrt{\pi}\Psi\left(\tau_j^*/\{1 + Q_{\bar{\Omega}}(\alpha_j^*)\}^{1/2}; \kappa_j^*/\{1 + Q_{\bar{\Omega}}(\alpha_j^*)\}, \nu\right)} \right\}^{1/\nu}$$

where $\alpha_j^* = \alpha_{\{j\}}^*$, $\tau_j^* = \tau_{\{j\}}^*$ *and* $\kappa_j^* = \kappa_{\{j\}}^*$ *are the marginal parameters (4.5) under Proposition 4.2(1). Then* $M_n/a_n \Rightarrow U$ *as* $n \to +\infty$, *where* U *has univariate* ν*-Fréchet marginal distributions (i.e.* $e^{-x^{-\nu}}$, $x > 0$*), and exponent function*

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu} \Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}} \left(\frac{x_i^+}{x_j^+} - \omega_{i,j} \right), i \in I_j \right)^\top ; \bar{\Omega}_j^+, \alpha_j^+, \tau_j^+, \nu + 1 \right), \quad (4.6)$$

where Ψ_{d-1} *is a* $(d-1)$*-dimensional central extended skew-t distribution with correlation matrix* $\bar{\Omega}_j^+$, *shape and extension parameters* α_j^+ *and* τ_j^+, *and* $\nu + 1$ *degrees of freedom,* $I = \{1, \ldots, d\}$, $I_j = I \backslash \{j\}$, *and* $\omega_{i,j}$ *is the* (i, j)*-th element of* $\bar{\Omega}$.

Proof (and further details) in Appendix D.1.4.

As the limiting distribution (4.6) is the same as that of the classic skew-t distribution (see Padoan, 2011), it exhibits identical upper and lower tail dependence coefficients (e.g. Joe, 1997, Ch 5). That is, the extension and non-centrality parameters, τ and κ, do not affect the extremal behavior.

4.2.2 A non-stationary, skew-normal random process

While there are several definitions of a stationary skew-normal process (e.g. Minozzo and Ferracuti, 2012), stationarity is incompatible with the requirement that all finite-dimensional distributions of the process are skew-normal. We now construct a non-stationary version of the skew-normal process through the additive-type stochastic representation (e.g. Azzalini, 2013, Ch. 5). A similar approach was explored by Zhang and El-Shaarawi (2010) for the stationary case.

Definition 4.4. Let $\{X(s)\}_{s\in\mathbb{S}}$ be a stationary Gaussian random process on $\mathbb{S}$ with zero mean, unit variance and correlation function $\rho(h) = \mathbb{E}\{X(s)X(s+h)\}$ for $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$. For $X' \sim \mathcal{N}(0,1)$ independent of $X(s)$, $\varepsilon \in \mathbb{R}$ and a function $\delta : \mathbb{S} \mapsto (-1,1)$, define

$$\begin{aligned} X''(s) &:= X'|X' + \varepsilon > 0, \qquad \forall\, s \in \mathbb{S} \\ Z(s) &:= \sqrt{1-\delta(s)^2}X(s) + \delta(s)X''(s), \quad s \in \mathbb{S}. \end{aligned} \tag{4.7}$$

Then $Z(s)$ is a skew-normal random process.

We refer to $\delta(s)$ as the slant function. From (4.7), if $\delta(s) \equiv 0$ for all $s \in \mathbb{S}$, then Z is a Gaussian random process. Note that Z is a random process with a consistent family of distribution functions, since $Z(s) = a(s)X(s) + b(s)Y(s)$ where a and b are bounded functions and X and Y are random processes with a consistent family of distribution functions. For any finite sequence of points $s_1, \ldots, s_d \in \mathbb{S}$ the joint distribution of $Z(s_1), \ldots, Z(s_d)$ is $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$, where

$$\begin{aligned} \bar{\Omega} &= D_\delta(\bar{\Sigma} + (D_\delta^{-1}\delta)(D_\delta^{-1}\delta)^\top)D_\delta, \\ \alpha &= \{1 + (D_\delta^{-1}\delta)^\top\bar{\Sigma}^{-1}(D_\delta^{-1}\delta)\}^{-1/2}D_\delta^{-1}\bar{\Sigma}^{-1}(D_\delta^{-1}\delta), \\ \tau &= \{1 + Q_{\bar{\Omega}}(\alpha)\}^{1/2}\,\varepsilon, \end{aligned} \tag{4.8}$$

and where $\bar{\Sigma}$ is the $d \times d$ correlation matrix of X, $\delta = (\delta(s_1), \ldots, \delta(s_d))^\top$ and $D_\delta = \{1_d - \mathrm{diag}(\delta^2)\}^{1/2}$, where 1_d is the identity matrix (Azzalini, 2013, Ch. 5). As a result, for any lag $h \in \mathbb{R}^k$, the distributions of $\{Z(s_1), \ldots, Z(s_d)\}$ and $\{Z(s_1+h), \ldots, Z(s_d+h)\}$ will differ unless $\delta(s) = 0$ for all $s \in \mathbb{S}$. Hence, the distribution of Z is not translation invariant and the process is not strictly stationary. For $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$, the mean $m(s)$ and covariance function $c_s(h)$ of the skew-normal random process are

$$m(s) = \mathbb{E}\{Z(s)\} = \delta(s)\phi(\varepsilon)/\Phi(\varepsilon)$$

and

$$c_s(h) = \mathrm{Cov}\{Z(s), Z(s+h)\} = \rho(h)\sqrt{\{1-\delta^2(s)\}\{1-\delta^2(s+h)\}} + \delta(s)\delta(s+h)(1-r), \tag{4.9}$$

where $r = \left\{\frac{\phi(\varepsilon)}{\Phi(\varepsilon)}\left(\varepsilon + \frac{\phi(\varepsilon)}{\Phi(\varepsilon)}\right)\right\}$. Hence, the mean is not constant and the covariance does not depend only on the lag h, unless $\delta(s) = \delta_0 \in (-1,1)$ for all $s \in \mathbb{S}$. In the latter case the skew-normal random process is weakly stationary (Zhang and El-Shaarawi, 2010).

One benefit of working with a skew-normal random field is that the non-stationary covariance function (4.9) is positive definite if the covariance function of X is positive definite, and if $-1 < \delta(s) < 1$ for all $s \in \mathbb{S}$. Hence, a valid model is directly obtainable by means of standard parametric correlation models $\rho(h)$ and any bounded function δ in $(-1,1)$. If the Gaussian process correlation function satisfies $\rho(0) = 1$ and $\rho(h) \to 0$ as $\|h\| \to +\infty$, then the correlation

of the skew-normal process is

$$\rho_s(h) = \frac{c_s(h)}{\sqrt{c_s(0)c_s(h)}} \approx \frac{\delta(s)\delta(s+h)(1-r)}{\sqrt{(1-\delta^2(s)r)(1-\delta^2(s+h)r)}},$$

as $\|h\| \to +\infty$, and $\rho_s(0) = 1$. Hence $\rho_s(h) = 0$ if either $\delta(s)$ or $\delta(s+h)$ are zero. Conversely, if both $\delta(s) \to \pm 1$ and $\delta(s+h) \to \pm 1$ then $\rho_s(h) \to \pm 1$.

The increments $Z(s+h) - Z(s)$ are skew-normal distributed for any $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$ (see Appendix A.4.5) and the variogram $2\gamma_s(h) = \mathrm{Var}\{Z(s+h) - Z(s)\}$ is equal to

$$2\gamma_s(h) = 2\left(1 - c_s(h) - \frac{\delta^2(s+h) + \delta^2(s)}{\pi}\right).$$

When $h = 0$ the variogram is zero, and when $\|h\| \to +\infty$ the variogram approaches a constant ≤ 2, respectively resulting in spatial independence or dependence for large distances h. We can now infer the conditions required so that $Z(s)$ has a continuous sample path.

Proposition 4.5. *Assume that $\mathbb{S} \subseteq \mathbb{R}$. A skew-normal process $\{Z(s), s \in \mathbb{S}\}$ has a continuous sample path if $\delta(s+h) - \delta(s) = o(1)$ and $1 - \rho(h) = O(|\log|h||^{-a})$ for some $a > 3$, as $h \to 0$.*

This result follows by noting that $r_s(h) = \rho(h) + \delta^2(s)(1-\rho(h)) + o(1)$ as $h \to 0$ and this is a consequence of the continuity assumption on $\delta(s)$, where $r_s(h) = c_s(h) + \{\delta^2(s+h) + \delta^2(s)\}/\pi$. Therefore, $1 - r_s(h) = O(|\log|h||^{-a})$ as $h \to 0$. Thus, the proof follows from the results in Lindgren (2012, page 48). This means that continuity of the skew-normal process is assured if $\delta(s)$ is a continuous function, in addition to the usual condition on the correlation function of the generating Gaussian process (e.g. Lindgren, 2012, Ch. 2).

Figure 4.1 illustrates trajectories of the skew-normal process for $k = 1$, with $X(s)$ a zero mean unit variance Gaussian process on $[0, 1]$ with isotropic power-exponential correlation function

$$\rho(h; \vartheta) = \exp\{-(h/\lambda)^\xi\}, \quad \vartheta = (\lambda, \xi),\ \lambda > 0,\ 0 < \xi \leq 2,\ h > 0, \tag{4.10}$$

with $\xi = 1.5$, $\lambda = 0.3$ and $h \in [0, 1]$. The first row shows the standard stationary case. The second row illustrates the non-stationary correlation function obtained with $s = 0.1$ (solid line), close to the stationary correlation. Thus the correlation of the process at two locations s and $s+h$ vanishes asymptotically as $|h| \to +\infty$. The second row illustrates the non-stationary correlation function obtained with $s = 0.1$ (solid line) behaving close to the stationary correlation, with this correlation decaying more slowly as s increases (but not approaching zero). The third row demonstrates both that points may be negatively correlated and that $\rho_s(h)$ is not necessarily a decreasing function in h. The bottom row highlights this even more clearly – correlation functions need not be monotonically decreasing – implying that pairs of points far apart can be more dependent than nearby points.

Simulating a skew-normal random process is computationally cheap through Definition 4.4, with the simulation of the required stationary Gaussian process achievable through many fast algorithms (e.g., Wood and Chan, 1994; Chan and Wood, 1997). Rather than relying on (4.8),

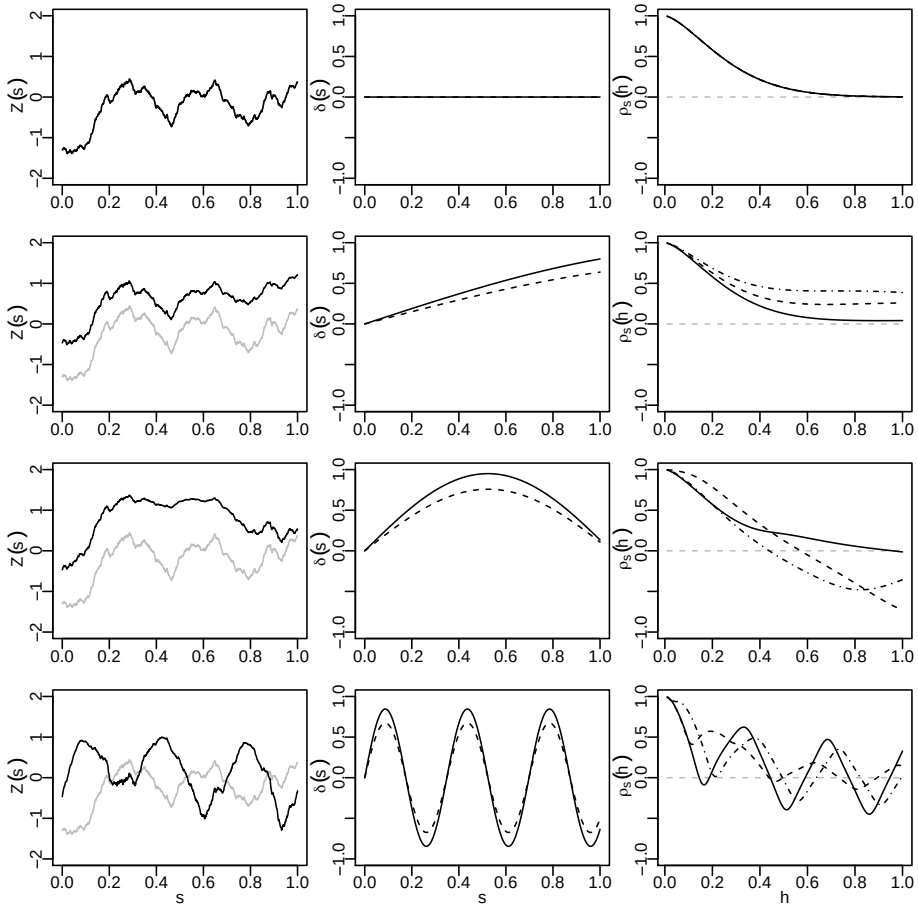

Figure 4.1: Simulations from four univariate skew-normal random processes on $[0, 1]$ with $\varepsilon = 0$. The left column shows the sample path (solid line) of the simulated process $Z(s)$ and of the generating Gaussian process $X(s)$ (grey line). The middle column illustrates the slant function $\delta(s)$ (solid line) and the mean $m(s)$ of the process (dashed line). The right column displays the non-stationary correlation functions at locations $s = 0.1$ (solid line), 0.5 and 0.75 (dot-dash). Rows 1–3 use slant function $\delta(s) = a\sin(bs)$ with $a = 0.95$ and $b = 0, 1$ and 3 respectively, whereas row 4 uses $\delta(s) = a^2\sin(bs)\cos(bs)$ with $a = 1.3$ and $b = 0.9$.

for practical purposes, to directly simulate from a skew-normal process with given parameters α, $\bar{\Omega}$ and τ, a conditioning sampling approach can be adopted (Azzalini, 2013, Ch. 5).

Specifically, let $X(s)$ define a zero-mean, unit variance stationary Gaussian random field on $\mathbb{S}$ with correlation function $\omega(h) = \mathbb{E}\{X(s)X(s+h)\}$ and let $\bar{\Omega}$ be the $d \times d$ correlation matrix of $X(s_1), \ldots, X(s_d)$. Specify $\alpha : \mathbb{S} \mapsto \mathbb{R}$ to be a continuous square-integrable function and let $\langle \alpha, X\rangle = \int_{\mathbb{S}} \alpha(s)X(s)\,\mathrm{d}s$ be the inner product. Let X' be a standard normal random variable independent of X and $\tau \in \mathbb{R}$. If we define

$$Z(s) = \left\{X(s) | \langle \alpha, X\rangle > X' - \tau\right\}, \qquad s \in \mathbb{S} \tag{4.11}$$

then, for any finite set $s_1, \ldots, s_d \in \mathbb{S}$, the distribution of $Z(s_1), \ldots, Z(s_d)$ is $\mathcal{SN}(\bar{\Omega}, \alpha, \tau)$, where $\alpha \equiv \{\alpha(s_1), \ldots, \alpha(s_d)\}$. For simplicity we also refer to $\alpha(s)$ as the slant function. More efficient simulation of skew-normal processes can be achieved by considering the form $Z(s) = X(s)$ if

$\langle \alpha, X \rangle > X' - \tau$ and $Z(s) = -X(s)$ otherwise (e.g. Azzalini, 2013, Ch. 5).

We conclude this section by discussing some extremal properties of the skew-normal process $Z(s)$. For a finite sequence of points $s_1, \ldots, s_d \in \mathbb{S}$, with $d \geq 2$. Each margin $Z(s_i)$ follows a skew-normal distribution (Azzalini, 2013) and so is in the domain of attraction of a Gumbel distribution (Chang and Genton, 2007; Padoan, 2011). Further, each pair $(Z(s_i), Z(s_j))$ is asymptotically independent (Bortot, 2010; Lysenko et al., 2009). However, in this case a broad class of tail behaviours can still be obtained by assuming that the joint survival function is regularly varying at $+\infty$ with index $-1/\eta$ (Ledford and Tawn, 1996), so that

$$\Pr(Z(s_i) > x, Z(s_j) > x) = x^{-1/\eta} \mathscr{L}(x), \qquad x \to +\infty, \tag{4.12}$$

where $\eta \in (0, 1]$ is the coefficient of tail dependence and $\mathscr{L}(x)$ is a slowly varying i.e., $\mathscr{L}(ax)/\mathscr{L}(x) \to 1$ as $x \to +\infty$, for fixed $a > 0$. Considering $\mathscr{L}$ as a constant, at extreme levels margins are negatively associated when $\eta < 1/2$, independent when $\eta = 1/2$ and positively associated when $1/2 < \eta < 1$. When $\eta = 1$ and $\mathscr{L}(x) \nrightarrow 0$ asymptotic dependence is obtained. We derive the asymptotic behavior of the joint survival function (4.12) for a pair of skew-normal margins. As our primary interest is in spatial applications, we focus on the joint upper tail of the skew-normal distribution when the variables are positively correlated or uncorrelated.

Proposition 4.6. *Let $Z \sim \mathcal{SN}_2(\bar{\Omega}, \alpha)$, where $\alpha = (\alpha_1, \alpha_2)^\top$ and $\bar{\Omega}$ is a correlation matrix with off-diagonal term $\omega \in [0, 1)$. The joint survivor function of the bivariate skew-normal distribution with unit Fréchet margins behaves asymptotically as* (4.12)*, where:*

1. *when either $\alpha_1, \alpha_2 \geq 0$, or $\omega > 0$ and $\alpha_j \leq 0$ and $\alpha_{3-j} \geq -\omega^{-1}\alpha_j$ for $j = 1, 2$, then*

$$\eta = (1 + \omega)/2, \qquad \mathscr{L}(x) = \tfrac{2\,(1+\omega)}{1-\omega} (4\pi \log x)^{-\omega/(1+\omega)};$$

2. *when $\omega > 0$, $\alpha_j < 0$, and $-\omega\,\alpha_j \leq \alpha_{3-j} < -\omega^{-1}\alpha_j$, for $j = 1, 2$, then*
   (a) *If $\alpha_{3-j} > -\alpha_j/\bar{\alpha}_j$ then*

$$\eta = \tfrac{(1-\omega^2)\bar{\alpha}_j^2}{1-\omega^2+(\bar{\alpha}_j-\omega)^2}, \qquad \mathscr{L}(x) = \tfrac{2\,\bar{\alpha}_j^2(1-\omega^2)}{(\bar{\alpha}_j^2-\omega)(1-\omega\bar{\alpha}_j)} (4\pi \log x)^{1/2\eta-1};$$

   (b) *If $\alpha_{3-j} < -\alpha_j/\bar{\alpha}_j$ then*

$$\eta = \left[\tfrac{1-\omega^2+(\bar{\alpha}_j-\omega)^2}{(1-\omega^2)\bar{\alpha}_j^2} + \left(\alpha_{3-j} + \tfrac{\alpha_j}{\bar{\alpha}_j}\right)^2 \right]^{-1},$$

$$\mathscr{L}(x) = \tfrac{-2^{3/2}\pi^{1/2}\bar{\alpha}_j^2(1-\omega^2)(\alpha_{3-j}+\alpha_j/\bar{\alpha}_j)^{-1}}{(\bar{\alpha}_j-\omega)\{1-\omega\bar{\alpha}_j+\alpha_j(\alpha_j+\alpha_{3-j}\bar{\alpha}_j)(1-\omega^2)\}} (4\pi \log x)^{1/2\eta-3/2};$$

3. *when either $\alpha_1, \alpha_2 < 0$, or $\omega > 0$, $\alpha_j < 0$ and $0 < \alpha_{3-j} < -\omega\,\alpha_j$ for $j = 1, 2$, then*

$$\eta = \left\{ \tfrac{1}{1-\omega^2} \left(\tfrac{\alpha_{3-j}^2(1-\omega^2)+1}{\bar{\alpha}_{3-j}^2} + \tfrac{\alpha_j^2(1-\omega^2)+1}{\bar{\alpha}_j^2} + \tfrac{2(\alpha_{3-j}\alpha_j(1-\omega^2)-\omega)}{\bar{\alpha}_{3-j}\bar{\alpha}_j} \right) \right\}^{-1},$$

$$\mathscr{L}(x) = \tfrac{-2^{3/2}\pi^{1/2}\bar{\alpha}_j^{3/2}\bar{\alpha}_{3-j}^2(1-\omega^2)(\alpha_i\bar{\alpha}_j+\alpha_j\bar{\alpha}_{3-j})^{-1}}{(\bar{\alpha}_j-\omega\bar{\alpha}_{3-j})\{1-\omega\bar{\alpha}_j+\alpha_j(\alpha_j+\alpha_{3-j}\bar{\alpha}_j/\bar{\alpha}_{3-j})(1-\omega^2)\}} (4\pi \log x)^{1/2\eta-3/2};$$

where $\bar{\alpha}_j = \sqrt{1+\alpha_j^{*2}}$ *and* $\alpha_j^* := \alpha_{\{j\}}^* = \frac{\alpha_j + \omega\alpha_{3-j}}{\sqrt{1+\alpha_{3-j}(1-\omega^2)}}$.

Proof in Appendix D.1.3.

As a result, when both marginal parameters are non-negative (case 1) then $1/2 \leq \eta < 1$, with $\eta = 1/2$ occurring when $\omega = 0$. As a consequence, as for the Gaussian distribution (for which $\alpha = 0$) the marginal extremes are either positively associated or exactly independent. The marginal extremes are also completely dependent when $\omega = 1$, regardless of the values of the slant parameters, α. When one marginal parameter is positive and one is negative (case 2) then $\eta > (1+\omega)/2$. In this case the extreme marginals are also positively associated, but the dependence is greater than when the random variables are normally distributed. Finally, when both marginal parameters are negative (case 3), then $0 < \eta < 1/2$ implying that the extreme marginals are negatively associated, although $\omega > 0$. It should be noted that differently from the Gaussian case ($\alpha = 0$) where $\omega > 0$ implies a positive association, in this case this is not necessary true. In summary, the degree of dependence in the upper tail of the skew-normal distribution ranges from negative to positive association and including independence.

4.3 Spectral representation for the extremal-skew-t process

The spectral representation of stationary max-stable processes with common unit Fréchet margins can be constructed using the fundamental procedures introduced by de Haan (1984) and Schlather (2002) (see also de Haan and Ferreira, 2006, Ch. 9). This representation can be formulated in broader terms resulting in max-stable processes with ν-Fréchet univariate marginal distributions, with $\nu > 0$ (Opitz, 2013). In order to state our result we rephrase the spectral representation so to also take into account non-stationary processes.

Let $\{Y(s)\}_{s\in\mathbb{S}}$ be a non-stationary real-valued stochastic process with continuous sample path on $\mathbb{S}$ such that $\mathbb{E}\{\sup_{s\in\mathbb{S}} Y(s)\} < \infty$ and $m^+(s) = \mathbb{E}[\{Y^+(s)\}^\nu] < \infty, \forall s \in \mathbb{S}$ for $\nu > 0$, where $Y^+(\cdot) = \max\{Y(\cdot), 0\}$ denotes the positive part of Y. Let $\{R_i\}_{i\geq 1}$ be the points of an inhomogeneous Poisson point process on $(0,\infty)$ with intensity $\nu r^{-(\nu+1)}$, $\nu > 0$, which are independent of Y. Define

$$U(s) = \max_{i=1,2,\ldots} \{R_i Y_i^+(s)\}/\{m^+(s)\}^{1/\nu}, \quad s \in \mathbb{S}, \tag{4.13}$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of Y. Then U is a max-stable process with common ν-Fréchet univariate margins. In particular, for fixed $s \in \mathbb{S}$ and $x(s) > 0$ we have

$$\Pr(U(s) \leq x(s)) = \exp\left[-\frac{\mathbb{E}\{Y^+(s)^\nu\}}{x^\nu(s) m^+(s)}\right] = \exp\{-1/x^\nu(s)\},$$

and for fixed $s_1, \ldots, s_d$ the finite dimensional distribution of U has exponent function

$$V(x(s_1), \ldots, x(s_d)) = \mathbb{E}\left(\max_j \left[\frac{\{Y^+(s_j)/x(s_j)\}^\nu}{m^+(s_j)}\right]\right), \quad x(s_j) > 0,\, j = 1, \ldots, d \tag{4.14}$$

(de Haan and Ferreira, 2006, Ch. 9).

In this construction, the impact of a non-stationary process $Y(s)$ would be that the dependence structure of the max-stable process $U(s+h)$ depends on both the separation h and the location $s \in \mathbb{S}$, and would therefore itself be non-stationary. The theorem below derives a max-stable process $U(s)$ when $Y(s)$ is the skew-normal random field introduced in Section 4.2.2.

Theorem 4.7 (Extremal skew-t process)**.** *Let $Y(s)$ be a skew-normal random field on $s \in \mathbb{S}$ with finite dimensional distribution $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$, as defined in equation (4.11). Then the max-stable process $U(s)$, given by (4.13), has ν-Fréchet univariate marginal distributions and exponent function*

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu} \Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^\circ}{x_j^\circ} - \omega_{i,j}\right), i \in I_j\right)^\top; \bar{\Omega}_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, \nu+1\right), \tag{4.15}$$

where $x_j \equiv x(s_j)$, Ψ_{d-1} is a $(d-1)$-dimensional non-central extended skew-t distribution (Definition 4.1) with correlation matrix $\bar{\Omega}_j^\circ$, shape, extension and non-centrality parameters $\alpha_j^\circ, \tau_j^\circ$ and κ_j°, $\nu+1$ degrees of freedom, $I = \{1, \ldots, d\}$, $I_j = I \backslash \{j\}$, and $\omega_{i,j}$ is the (i,j)-th element of $\bar{\Omega}$.

Proof (and further details) in Appendix D.1.5.

We call the process $U(s)$ with exponent function (4.15) an extremal skew-t process.

Note that in Theorem 4.7 when $\tau = 0$, and the slant function is such that $\alpha(s) \equiv 0$ for all $s \in \mathbb{S}$, then the exponent function (4.15) becomes

$$V(x_j, j \in I) = \sum_{j \in I} x_j^{-\nu} \Psi_{d-1}\left[\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i}{x_j} - \omega_{i,j}\right), i \in I_j\right)^\top; \bar{\Omega}_j^\circ, \nu+1\right]. \tag{4.16}$$

This is the exponent function of the extremal-t process as discussed in Opitz (2013).

If we assume $\tau = 0$ in (4.11), then the bivariate exponent function of the extremal skew-t process seen as a function of the separation h is equal to

$$V\{x(s), x(s+h)\} = \frac{\Psi(b(x_s^*(h)); \alpha_s^*(h), \tau_s^*(h), \nu+1)}{x^\nu(s)} + \frac{\Psi(b(x_s^+(h)); \alpha_s^+(h), \tau_s^+(h), \nu+1)}{x^\nu(s+h)}$$

where Ψ is a univariate extended skew-t distribution, $b(\cdot) = \sqrt{\frac{\nu+1}{1-\omega^2(h)}}(\cdot - \omega(h))$,

$$\begin{aligned}
x_s^*(h) &= \tfrac{x(s+h)\Gamma_s(h)}{x(s)}, & x_s^+(h) &= \tfrac{x(s)}{x(s+h)\Gamma_s(h)},\\
\alpha_s^*(h) &= \alpha(s+h)\sqrt{1-\omega^2(h)}, & \alpha_s^+(h) &= \alpha(s)\sqrt{1-\omega^2(h)},\\
\tau_s^*(h) &= \sqrt{\nu+1}\{\alpha(s) + \alpha(s+h)\omega(h)\}, & \tau_s^+(h) &= \sqrt{\nu+1}\{\alpha(s+h) + \alpha(s)\omega(h)\},
\end{aligned}$$

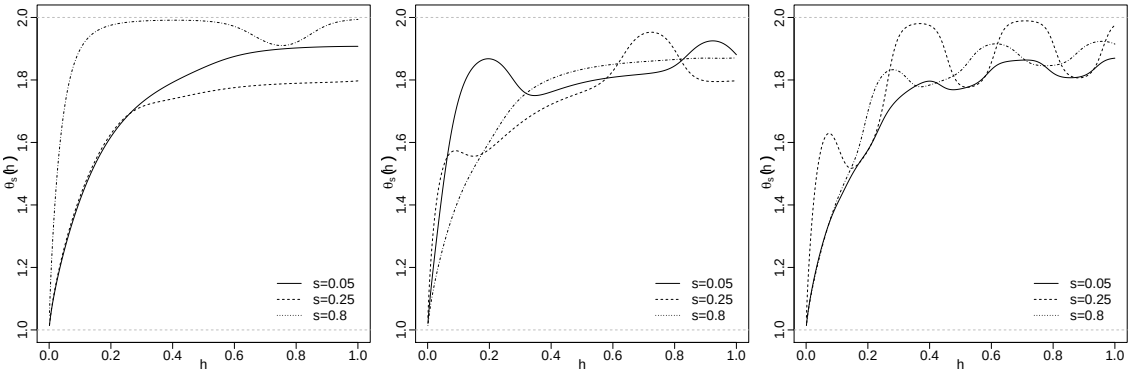

Figure 4.2: Examples of univariate ($k = 1$) non-stationary isotropic extremal coefficient functions $\theta_s(h)$, for the extremal skew-t process over $s \in [0, 1]$, using correlation function (4.10) where $h \in [0, 1]$, $\lambda = 1.5$ and $\gamma = 0.3$. Slant functions are (left to right panels): $\alpha(s) = -1 - s + \exp\{\sin(5s)\}, \alpha(s) = 1 + 1.5s - \exp\{\sin(8s)\}$ and $\alpha(s) = 2.25\sin(9s)\cos(9s)$. Solid, dashed and dot-dashed lines represent the fixed locations $s = 0.05, 0.25$ and 0.8 respectively.

and

$$\Gamma_s(h) = \left(\frac{\Psi\left[\alpha(s) + \alpha(s+h)\omega(h)\sqrt{\frac{\nu+1}{\alpha^2(s+h)\{1-\omega^2(h)\}}}; \nu+1\right]}{\Psi\left[\alpha(s+h) + \alpha(s)\omega(h)\sqrt{\frac{\nu+1}{\alpha^2(s)\{1-\omega^2(h)\}}}; \nu+1\right]} \right)^{1/\nu}.$$

Clearly, as the dependence structure depends on both the correlation function $\omega(h)$ and the slant function $\alpha(s)$, and therefore on the value of $s \in \mathbb{S}$, it is non-stationary dependence. From the bivariate exponent function we can derive the non-stationary extremal coefficient function, using the relation $\theta_s(h) = V(1, 1)$, which gives

$$\theta_s(h) = \Psi(b(\Gamma_s(h)); \alpha_s^*(h), \tau_s^*(h), \nu+1) + \Psi(b(1/\Gamma_s(h)); \alpha_s^+(h), \tau_s^+(h), \nu+1). \tag{4.17}$$

Figure 4.2 shows some examples of univariate ($k = 1$) non-stationary isotropic extremal coefficient functions obtained from (4.17) using the power-exponential correlation function (4.10). Each panel illustrates a different slant function $\alpha(s)$, with the line-types indicating the fixed location value of $s \in \mathbb{S}$. The extremal coefficient functions $\theta_s(h)$ increase as the value of h increases, meaning that the dependence of extremes decreases with the distance. $\theta_s(h)$ grows with different rates depending on the location $s \in \mathbb{S}$. Although the ergodicity and mixing properties of the process must be investigated, numerical results show that for some s, $\theta_s(h) \to 2$ as $|h| \to +\infty$. By increasing the complexity of the slant function (e.g. centre and right panels) it is possible to construct extremal coefficient functions which exhibit stronger dependence for larger distances, h, compared to shorter distances. Similarly Figure 4.3 illustrates examples of bivariate ($k = 2$) non-stationary geometric anisotropic extremal coefficient functions, $\theta_s(h)$, also obtained from (4.17). Similar interpretations to the univariate case can be made (Figure 4.2), in addition to noting that the level of dependence is affected by the direction (from the origin).

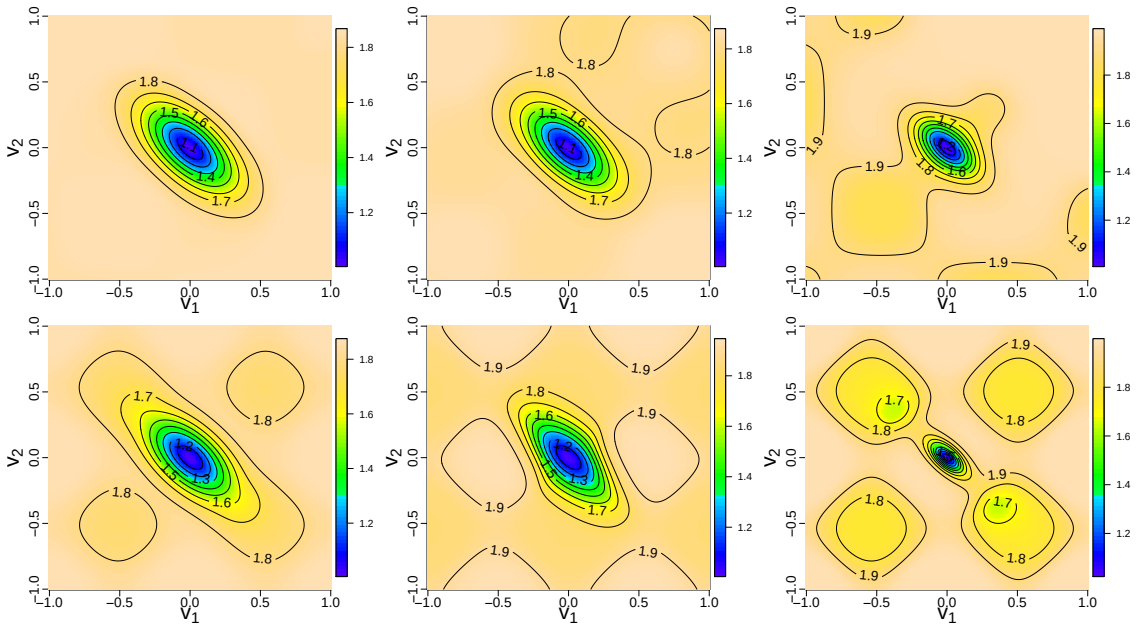

Figure 4.3: Bivariate ($k = 2$) geometric anisotropic non-stationary extremal coefficient functions $\theta_s(h)$, for the extremal skew-t process on $s \in [0,1]^2$, based on extremal coefficient function (4.17) with $\lambda = 1.5$ and $\gamma = 0.3$, where $h = v^\top R v$, $v = (v_1, v_2)^\top \in [-1,1]^2$ and R is a 2×2 matrix whose diagonal elements are 2.5 and off-diagonal elements 1.5. Slant functions are $\alpha(s) = \exp\{\sin(4s_1)\sin(4s_2) - s_1 s_2 - 1\}$ (top panels) and $\alpha(s) = 2.25\{\sin(3s_1)\cos(3s_1) + \sin(3s_2)\cos(3s_2)\}$ (bottom), with $s = (s_1, s_2)^\top \in [0,1]^2$. Left to right, panels are based on fixing $s = (0.2, 0.2)^\top$, $s = (0.4, 0.4)^\top$ and $s = (0.85, 0.85)^\top$ (top panels) and $s = (0.25, 0.25)^\top$, $s = (0.25, 0.8)^\top$ and $s = (0.8, 0.8)^\top$ (bottom).

4.4 Inference for extremal skew-t processes

Parametric inference for the extremal-skew-t process can be performed in two ways. The first uses the marginal composite-likelihood approach (e.g. Padoan et al., 2010; Davison and Gholamrezaee, 2012; Huser and Davison, 2013), since only marginal densities of dimension up to $d = 4$ are practically available (see Appendix D.2.1).

Let $\vartheta \in \Theta$ denote the vector of dependence parameters of the extremal-skew-t process. Consider a sample $x_1, \ldots, x_n \in \mathbb{R}_+^k$ of n i.i.d. replicates of the process observed over a finite number of points $s_1, \ldots, s_d \in \mathbb{S}$. For simplicity, it is assumed that the univariate marginal distributions are unit Fréchet. The pairwise or triplewise ($m = 2, 3$) log-composite-likelihood is defined by

$$\ell_m(\vartheta; x) = \sum_{i=1}^{n} \sum_{E \in E_m} \log f(x_i \in E; \vartheta), \quad m = 2, 3,$$

where $x = (x_1, \ldots, x_n)^\top$ and f is a marginal extremal-skew-t pdf associated with each member of a set of marginal events E_m. See e.g. Varin et al. (2011) for a complete description of composite likelihood methods.

A second approach is to use the approximate likelihood function introduced by Coles and Tawn (1994), which is constructed on the space of angular densities. The angular measure of the extremal-skew-t dependence model (4.15) places mass on the interior as well as on all the

other subspaces of the simplex, such as the edges and the vertices. We derive some of these densities following the results in Coles and Tawn (1991).

Let J be an index set that takes values in $\mathbb{I} = \mathbb{P}(\{1,\dots,d\})\backslash\emptyset$, where $\mathbb{P}(I)$ is the power set of I. For any fixed d and all $J \in \mathbb{I}$, the sets

$$\mathbb{W}_{d,J} = (w \in \mathbb{W} : w_j = 0, \text{ if } j \notin J;\ w_j > 0 \text{ if } j \in J)$$

provide a partition of the d-dimensional simplex $\mathbb{W}$ into $2^d - 1$ subsets. Let $k = |J|$ be the size of J. Let $h_{d,J}$ denote the density that lies on the subspace $\mathbb{W}_{d,J}$, which has $k-1$ free parameters w_j such that $j \in J$. When $J = \{1,\dots,d\}$ the angular density in the interior of the simplex is

$$h(w) = \frac{\psi_{d-1}\left(\left[\sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}}\left\{\left(\frac{w_i^\circ}{w_1^\circ}\right)^{1/\nu} - \omega_{i,1}\right\}, i \in I_1\right]^\top ; \Omega_1^\circ, \alpha_1^\circ, \tau_1^\circ, \kappa_1^\circ, \nu+1\right)}{w_1^{(d+1)}\left\{\prod_{i=2}^d \frac{1}{\nu}\sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}}\left(\frac{w_i^\circ}{w_1^\circ}\right)^{\frac{1}{\nu}-1}\frac{m_i^+}{m_1^+}\right\}^{-1}}, \quad w \in \mathbb{W} \tag{4.18}$$

where ψ_{d-1} denotes the $d-1$-dimensional skew-t density, $I_j = \{1,\dots,d\}\backslash j$ and where the parameters $\Omega_1^\circ, \alpha_1^\circ, \tau_1^\circ, \kappa_1^\circ$ and $w_i^\circ = x_i(m_i^+)^{1/\nu}$ are given in the proof to Theorem 4.7 (Appendix D.1.5). When $J = \{i_1,\dots,i_k\} \subset \{1,\dots,d\}$, the angular density for any $x \in \mathbb{R}_+^d$ is

$$h_{d,J}\left(\frac{x_{i_1}}{\sum_{i\in J} x_i}, \cdots, \frac{x_{i_{k-1}}}{\sum_{i\in J} x_i}\right) = -\left(\sum_{i\in J} x_i\right)^{k+1} \lim_{\substack{x_j\to 0,\\ j\notin J}} \frac{\partial^k V}{\partial x_{i_1}\cdots\partial x_{i_k}}(x). \tag{4.19}$$

Thus, when $J = \{j\}$ for any $j \in \{1,\dots,d\}$ then $\mathbb{W}_{d,J}$ is a vertex $\boldsymbol{e}_j$ of the simplex and the density is a point mass, denoted $h_{d,J} = H(\{\boldsymbol{e}_j\})$. In this case (4.19) reduces to

$$h_{d,J} = \Psi_{d-1}\left\{\left(-\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\omega_{i,j}, i \in I_j\right)^\top ; \Omega_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, \nu+1\right\}, \tag{4.20}$$

where Ψ_{d-1} denotes the $d-1$-dimensional skew-t distribution with parameters again given in the proof to Theorem 4.7 (Appendix D.1.5).

Computations of all $2^d - 1$ densities that lie on the edges and vertices of the simplex are available for $d = 3$. In this case, the angular densities on the interior and vertices of the simplex can be deduced from (4.18) and (4.20). For all $i, j \in J = \{1,2,3\}$, with $i \neq j$, the angular

density on the edges of $\mathbb{W}_{d,J}$ for $w \in \mathbb{W}_{d,J}$ is given by

$$
\begin{aligned}
h_{3,\{i,j\}}(w) = \sum_{u,v\in\{i,j\},u\neq v} \Bigg(& \frac{\psi(b^{\circ}_{u,v};\nu+1)}{\Psi(\bar{\tau}_u;\nu+1)} \Psi_2\left[\{y^{\circ}_1(u,v), y^{\circ}_2(u,v)\}^\top; \bar{\Omega}^{\circ\circ}_u, \nu+2\right] \\
& \times \frac{1}{w_1}\left\{\frac{\mathrm{d}^2 b^{\circ}_{u,v}}{\mathrm{d}w_u \mathrm{d}w_v} + \frac{\mathrm{d}b^{\circ}_{u,v})}{\mathrm{d}w_v}\left(\frac{\mathrm{d}b^{\circ}_{u,v}}{\mathrm{d}w_u}\frac{(\nu+2)b^{\circ}_{u,v}}{\nu+1+b^{\circ 2}_{u,v}} - \frac{1}{w_1}\right)\right\} \\
& + \psi\{y^{\circ}_1(u,v);\nu+2\}\sqrt{\frac{\nu+2}{1-\Omega^{\circ 2}_{u,[1,2]}}}\frac{b^{\circ}_{u,v}c_{u,\bar{k}} + \Omega^{\circ 2}_{u,[1,2]}(\nu+1)}{(\nu+1+b^{\circ 2}_{u,v})^{3/2}} \\
& \times \Psi\left(\frac{\sqrt{\nu+3}\left\{z^{\circ}_2(u,v)\Omega^{\circ\circ}_{u,[1,1]} - z^{\circ}_1(u,v)\Omega^{\circ\circ}_{u,[1,2]}\right\}}{\sqrt{\left[\Omega^{\circ\circ}_{u,[1,1]}\{\nu+1+b^{\circ 2}_{u,v}\} + z^{\circ 2}_1(u,v)\right]\det(\Omega^{\circ\circ}_u)}}; \nu+3\right) \\
& + \psi\{y^{\circ}_2(u,v);\nu+2\}\sqrt{\frac{\nu+2}{1-\Omega^{*2}_{u,[1,3]}}}\frac{x(u,v)\bar{\tau}_u + \Omega^{*2}_{u,[1,3]}(\nu+1)}{\{\nu+1+b^{\circ 2}_{u,v}\}^{3/2}} \\
& \times \Psi\left(\frac{\sqrt{\nu+3}\left\{z^{\circ}_1(u,v)\Omega^{\circ\circ}_{u,[2,2]} - z^{\circ}_2(u,v)\Omega^{\circ\circ}_{u,[1,2]}\right\}}{\sqrt{\left(\Omega^{\circ\circ}_{u,[2,2]}\{\nu+1+b^{\circ 2}_{u,v}\} + z^{\circ 2}_2(u,v)\right)\det(\Omega^{\circ\circ}_u)}}; \nu+3\right)\Bigg),
\end{aligned}
\tag{4.21}
$$

where for all $u, v \in J$, with $u \neq v$, and $\bar{k} \notin \{i,j\}$,

$$
y^{\circ}_{\ell}(u,v) = \frac{z^{\circ}_{\ell}(u,v)}{\sqrt{\Omega^{\circ}_{u,[\ell,\ell]}}}\sqrt{\frac{\nu+2}{\nu+1+b^{\circ 2}_{u,v}}},\ \ell = 1,2, \quad z^{\circ}_1(u,v) = c_{u,\bar{k}} - \Omega^{\circ}_{u,[1,2]}b^{\circ}_{u,v},
$$

$$
c_{u,v} = -\omega_{u,v}\sqrt{\frac{\nu+1}{1-\omega^2_{u,v}}}, \quad z^{\circ}_2(u,v) = \bar{\tau}_u - \Omega^{\circ}_{u,[1,3]}, \quad b^{\circ}_{u,v}, \quad \Omega^{\circ}_u = \begin{bmatrix} \bar{\Omega}_u & -\delta_u \\ -\delta^\top_u & 1 \end{bmatrix},
$$

$$
\delta^\top_u = \bar{\Omega}_u\left(\alpha_v\sqrt{1-\omega^2_{u,v}}, \alpha_k\sqrt{1-\omega^2_{u,k}}\right)^\top,\ \bar{\Omega}^{\circ\circ}_u = \omega^{\circ -1/2}_u \Omega^{\circ\circ}_u \omega^{\circ -1/2}_u,
$$

$\omega^{\circ}_u = \mathrm{diag}(\Omega^{\circ\circ}_u)$, $\Omega^{\circ\circ}_u = \Omega^{\circ}_{u,[-1,-1]} - \Omega^{\circ}_{u,[-1,1]}\Omega^{\circ}_{u,[1,-1]}$. Components of Ω°_u and $\Omega^{\circ\circ}_u$ are respectively given by $\Omega^{\circ}_{u,[i,j]}$ and $\Omega^{\circ\circ}_{u,[i,j]}$ for $i,j \in J$. See also Appendix D.1.5 for further details. When, $\tau = 0$ and $\alpha(s) = 0$, then the densities (4.18), (4.20) and (4.21) reduce to the densities of the extremal-t dependence model.

Figure 4.4 illustrates some examples of the flexibility of the trivariate extremal-skew-t dependence structure. Here we write the correlation coefficients as $\omega = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3})^\top$ and the slant parameters as $\alpha = (\alpha_{1,2}, \alpha_{1,3}, \alpha_{2,3})^\top$, and assume that $\nu = 2$ and $\tau = 0$ for simplicity.

The plots in the left column have $\alpha = (0,0,0)^\top$ and so correspond to the extremal-t angular measure. The density in the top-left panel, obtained with $\omega = (0.52, 0.71, 0.52)^\top$, has mass concentrations mainly on the edge that links the first and the third variable, since they are the most dependent ($w_{1,3} = 0.71$). Some mass is also placed on the corners of the second variable, indicating that this is less dependent on the others ($w_{1,2} = w_{2,3} = 0.52$), and on the middle of the simplex, because a low degree of freedom ($\nu = 2$) pushes mass towards the centre of the simplex. The top-middle and top-right panels are extremal skew-t angular densities obtained

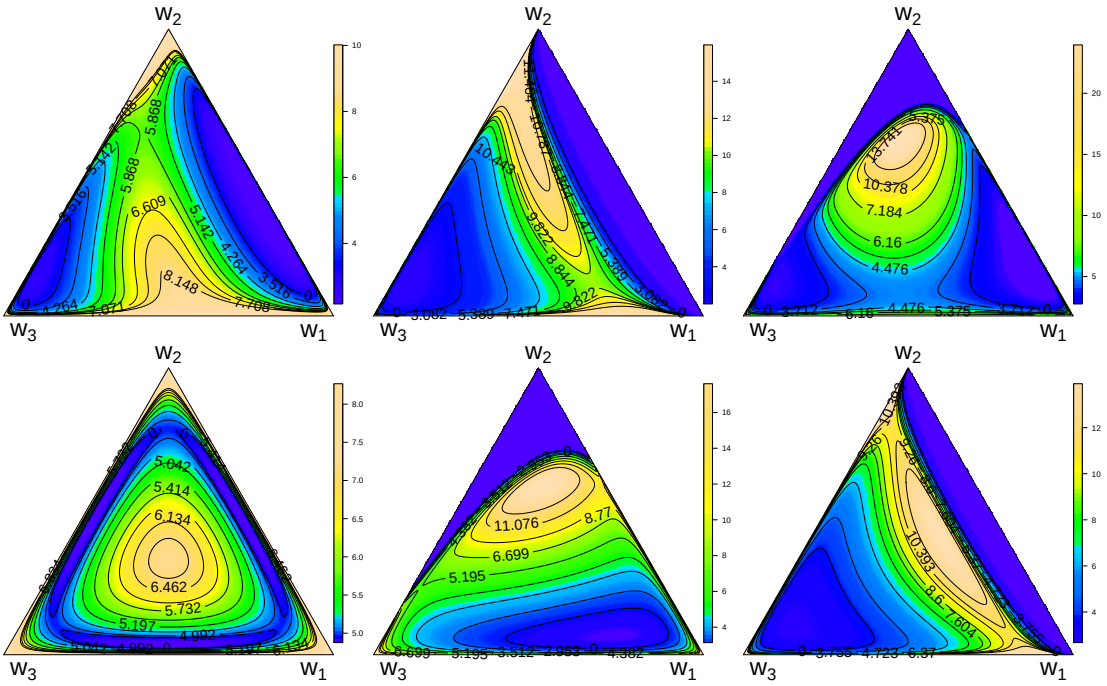

Figure 4.4: Trivariate extremal skew-t angular densities with $\nu = 2$ degrees of freedom. Correlation coefficients are $\omega = (0.52, 0.71, 0.52)^\top$ for the top row and $\omega = (0.95, 0.95, 0.95)^\top$ for the bottom. Left and centre columns respectively have skewness $\alpha = (0, 0, 0)^\top$ and $\alpha = (-2, -2, 5)^\top$. Right column has $\alpha = (-10, 5, 1)^\top$ (top) and $\alpha = (-2, -2, -2)^\top$ (bottom). In all cases $\tau = 0$ for simplicity.

with $\alpha = (-2, -2, 5)^\top$ and $\alpha = (-10, 5, 1)^\top$ respectively. Here the impact of the slant parameter is to increase the levels of dependence – indeed the mass is clearly pushed towards the centre of the simplex. In the middle panel dependence between the second and third variables has increased, while in the right panel all variables are strongly dependent with a greater dependence of the second variable on the others.

The bottom row in Figure 4.4 illustrates the spectral densities with correlation coefficients $\omega = (0.95, 0.95, 0.95)^\top$. The bottom-left panel is the standard extremal-t dependence (with $\alpha = (0, 0, 0)^\top$), which has a symmetric density with mass concentrated mainly in the centre of the simplex and on the vertices. The top-middle and top-right panels show extremal skew-t densities, obtained with $\alpha = (-2, -2, 5)^\top$ and $\alpha = (-2, -2, -2)^\top$ respectively. In this case the impact of the slant parameter is to decrease the dependence – here the mass is pushed towards the edges of the simplex. In the middle panel the second variable has become more dependent on the others, while in the right panel the third variable is more dependent on the others. These examples illustrate the great flexibility of the extremal skew-t model in capturing a wide range of extremal dependence behaviour above and beyond that of the standard extremal t model.

Therefore, for $d = 3$ the estimation of dependence parameters can be based on the following approach. Let $\{(r_i, w_i) : i = 1, \ldots, n\}$ be the set of observations, where $r_i = x_{i,1} + \cdots + x_{i,d}$ and $w_i = x_i / r_i$ are pseudo-polar radial and angular components. Then the approximate log-

likelihood is defined by

$$\ell(\vartheta; w) = \sum_{\substack{i=1,\dots,n: \\ r_i > r_0}} \log h(w_i; \vartheta), \tag{4.22}$$

where $w = (w_1, \dots, w_n)^\top$, for some radial threshold $r_0 > 0$, and where h is the angular density function of the extremal-skew-t dependence model. The components of the sum in (4.22) comprise the three types of angular densities lying on the interior, edges and vertices of the simplex. Whether an angular component belongs either to the interior, an edge or a vertex of the simplex, producing the associated density, is determined according the following criterion. We select a threshold $c \in [0, 0.1]$ and we construct the following partitions for an arbitrary (w_i, w_j, w_k): When $\mathcal{C}_i = \{w_i > 1 - c, w_j < c, w_k < c\}$ for $i = 1, 2, 3$, $j \neq k \neq i$ then an observation belongs to vertex $\boldsymbol{e}_i$. When $\mathcal{E}_{ij} = \{w_i, w_j < 1 - c, w_k < c, w_i > 1 - 2w_j, w_j > 1 - 2w_i\}$, for $i, j = 1, 2, 3$, $i \neq j$, $k \neq i, j$, then an observation belong to edge between the ith and jth components. When $\mathcal{I} = \{w_i < c, i = 1, 2, 3\}$ then an observation belongs to the interior. Figure 4.5 displays with the gray and shaded areas the partitions of the three-dimensional simplex into three corners and edges respectively. Observations which angular components fall into such areas are considered to belong to the corresponding subset of the simplex (corner, edge or interior).

For example, when $w_3 > 1-c$ (on the left of the green dashed line indicating the $1-c$ level for w_3), then $w = (w_1, w_2, w_3)$ is in the corner associated to the third component which corresponds to the grey shaded triangle on the bottom left of the simplex. Similarly, if both w_1 and w_2 are less than $1 - c$ (i.e. to the left of the blue dashed line indicating the $1 - c$ level of w_1 and below the red dashed line indicating the $1 - c$ level of w_2), such that $w1 > 1 - 2w_2$ and $w2 > 1 - 2w_1$ (i.e. to the right of the black dashed line bisecting the corner of the second component and above the black dashed line bisecting the corner of the first component) and if $w_3 < c$ (to the right of the green dashed line indicating the $1 - c$ level of w_3) then $w = (w_1, w_2, w_3)$ is on the edge

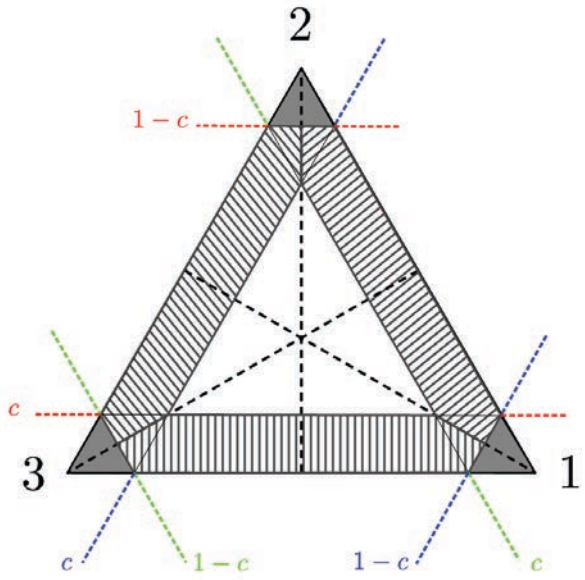

Figure 4.5: Partitions of the three-dimensional simplex

between the first and second component which correspond to the positively hatched area on the right hand side of the simplex. Finally if $w_1, w_2, w_3 > c$ (i.e. to the right of the blue dashed line, above the red dashed line and to the left of the green dashed line, respectively indicating the c levels of w_1, w_2 and w_3) then $w = (w_1, w_2, w_3)$ is in the interior of the simplex, represented by the white triangle in the centre of the simplex.

Observations then need to be rescaled so that the total mass is equal to the number of components. That is

$$\int_{\mathbb{W}} h(w)\mathrm{d}w = k_{\mathcal{C}_i} \sum_{i=1}^{3} \int_{\mathcal{C}} h_{3,\{i\}}\mathrm{d}w + k_{\mathcal{E}_{ij}} \sum_{i=1}^{3} \int_{\mathcal{E}_{ij}} h_{2,\{i,j\}}(w)\mathrm{d}w + k_{\mathcal{I}} \int_{\mathcal{I}} h_{3,\{1,2,3\}}(w)\mathrm{d}w = 3,$$

where

$$k_{\mathcal{C}_i} = \frac{4}{\sqrt{3}c^2}, \quad k_{\mathcal{E}_{ij}} = \frac{\int_0^1 h_{3,\{i,j\}}}{\frac{c}{2}\sqrt{3}(1-2c)}, \quad k_{\mathcal{I}} = \frac{\int_0^1 \int_0^1 h_{3,\{1,2,3\}}(w)\mathrm{d}w}{\int_c^{1-2c} \int_c^{1-2c} h_{3,\{1,2,3\}}(w)\mathrm{d}w}.$$

In the bivariate case ($d = 2$), the appropriate modification only considers the mass on the vertices and interior.

We now illustrate the ability of the approximate likelihood in estimating the extremal dependence parameters in the bivariate and trivariate cases. We generate 500 replicate datasets of sizes 5000 (bivariate) and 1000 (trivariate), with parameters $\vartheta_2 = (\omega, \nu) = (0.6, 1.5)$ and $\vartheta_3 = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3}, \nu) = (0.6, 0.8, 0.7, 1)$. Each dataset is transformed to pseudo-polar coordinates and the 100 observations with the largest radial component are retained. Parameters are estimated through the profile likelihood where the dependence parameter ω is the parameter of interest and the degree of freedom ν is considered as a nuisance parameter. Parameters are estimated for different values of the threshold $c = 0, 0.02, 0.04, 0.06, 0.08, 0.1$. In order to compare likelihoods for different values of c, the likelihood functions are evaluated using those data points considered to belong to the interior of the simplex, multiplied by the mass at the corners and/or edges in proportion to their rescaling constants.

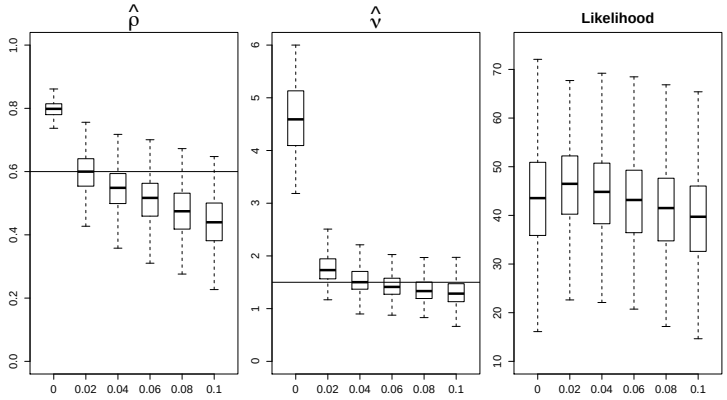

Figure 4.6: Left to right: Boxplots of the estimates of the dependence parameter ω, the degree of freedom ν and the associated maximum of the likelihood function based on the rescaled angular density, when $c = 0, 0.02, 0.04, 0.06, 0.08$ and 0.1. Boxplots are constructed from 500 replicate datasets of size 5000. Horizontal lines indicate the true values $\omega = 0.6$ and $\nu = 1.5$.

Figures 4.6 and 4.7 provide (left to right) boxplots of the resulting estimates of the dependence parameter(s) ω, the degree of freedom ν and of the likelihood function for increasing values of c, for the 500 replicate datasets for both bivariate and trivariate cases. The true parameter values are indicated by the horizontal lines.

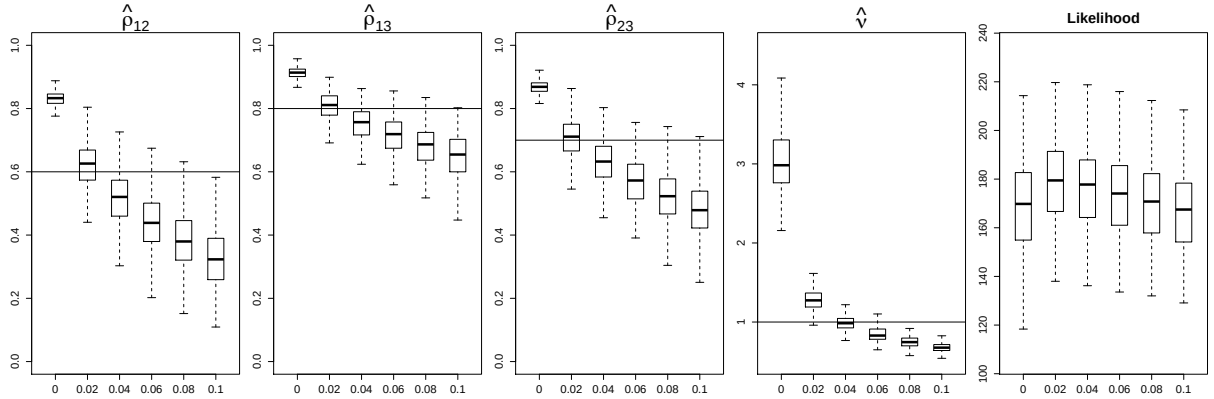

Figure 4.7: Left to right: Boxplots of the estimates of the dependence parameter $\omega = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3})$, the degree of freedom ν and the associated maximum of the likelihood function based on the rescaled angular density, when $c = 0, 0.02, 0.04, 0.06, 0.08$ and 0.1. Boxplots are constructed from 500 replicate datasets of size 1000. Horizontal lines indicate the true values $\omega_{1,2} = 0.6, \omega_{1,3} = 0.7, \omega_{2,3} = 0.7$ and $\nu = 1$.

In the rightmost panel of each Figure, the largest values of the log-likelihood are globally obtained for $c = 0.02$, for which the most accurate estimates of ω and ν are also obtained. Conditional on $c = 0.02$ the mean estimates are $\hat{\omega} = 0.55$ and $\hat{\nu} = 1.79$ in the bivariate case and $\hat{\omega} = (0.62, 0.80, 0.71)$ and $\hat{\nu} = 1.27$ in the trivariate case. Note that the degree of freedom ν appears to be slightly overestimated, and appears to be better estimated for slightly larger values of c. Overall this procedure appears capable of efficiently estimating the model parameters. Note that increased precision of estimates can be obtained by considering a denser range of threshold values c.

An independent study comparing the efficiency of the maximum pairwise and triplewise composite likelihood estimators is provided in Appendix D.3. This simulation study related to that of Huser and Davison (2013), demonstrates that the triplewise maximum-likelihood estimator can be more efficient than the pairwise maximum-likelihood estimator for the extremal-t process.

4.5 Application to wind speed data

We illustrate use of the extremal skew-t process using wind speed data (the weekly maximum wind speed in km/h), collected from 4 monitoring stations across Oklahoma, USA, over the March-May period during 1996–2012, as part of a larger dataset of 99 stations. An analysis establishing the marginal, station-specific skewness of these data is presented below. The maximum daily observations of wind speed (1564 observations per station) are considered. The t and skew-t distributions are fitted to the data using the maximum likelihood approach and a chi-square test is performed in order to investigate if the slant parameter of the skew-t distribu-

tion is significantly different from zero. Additionally the Fisher-Pearson coefficient of skewness (γ) is calculated.

The marginal estimation results are collected in Table 4.1. The estimated parameters are location μ, scale σ and degrees of freedom ν for the t distribution and in addition also the slant α for the skew-t distributions. The table also displays the p-value of a chi-square test for the 4 monitoring stations (CLOU, CLAY, SALL and PAUL) that exhibit the most skewness according to the slant coefficient. With a p-value of zero, the skewness of the data is proved marginally.

The red and blue solid lines in Figure 4.8 respectively show the fitted t and skew-t densities compared to the histogram of the daily observations for each of the 4 monitoring stations highlighted above. Each of the plots clearly shows that the datasets are right skewed and that the model with the ability to handle skewness provides a better fit.

Now we focus on the dependence structure between stations, where for simplicity the data is marginally transformed to unit Fréchet distributions. Only extremal-t and extremal skew-t models are considered, and parameter estimation is performed via pairwise composite likelihoods detailed at the beginning of Section 4.4.

Model comparison is performed through the composite likelihood information criterion (CLIC; Varin et al. (2011)) given by

$$\text{CLIC} = -2\left[\ell(\hat{\vartheta}) - \text{tr}\{\hat{J}(\hat{\vartheta})\hat{H}(\hat{\vartheta})^{-1}\}\right],$$

where $\hat{\vartheta}$ are the maximum composite likelihood estimates, $\ell(\hat{\vartheta})$ is the maximised pairwise composite likelihood, and $\hat{J}$ and $\hat{H}$ are estimates of $J(\vartheta) = \text{Var}(\nabla\ell(\vartheta))$ and $J(\vartheta) = \mathbb{E}(-\nabla^2\ell(\vartheta))$, the variability and sensibility (hessian) matrices.

Table 4.2 presents the pairwise composite likelihood estimates of $\omega = (\omega_{12}, \omega_{13}, \omega_{23})$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and ν for the extremal-t and extremal skew-t models obtained for all triplewise combinations of the four locations CLOU, CLAY, PAUL and SALL. For each triple the extremal skew-t model achieves a lower CLIC score than the extremal-t model, indicating its greater suitability. Moreover the standard errors of the slant parameters α, that are $(0.04, 0.14, 0.03)$, $(0.05, 0.97, 1.09)$, $(0.17, 0.15, 0.63)$ and $(0.03, 0.02, 3.49)$ for the four triples on Table 4.2 from top to the bottom respectively, prove that these parameters are non-zero, strengthening the

Station	Model	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\nu}$	p-value	γ
CLOU	t	11.84	2.75	–	5.78	–	–
	skew-t	8.51	20.24	2.79	11.21	0	1.17
CLAY	t	12.63	3.50	–	6.40	–	–
	skew-t	8.23	35.53	3.28	16.61	0	1.12
SALL	t	14.66	4.27	–	7.47	–	–
	skew-t	9.02	58.76	4.20	50.98	0	0.92
PAUL	t	15.76	4.25	–	9.31	–	–
	skew-t	11.43	38.55	1.78	17.81	0	0.79

Table 4.1: Outcome of the marginal analysis of the 4 stations that exhibit the strongest skewness.

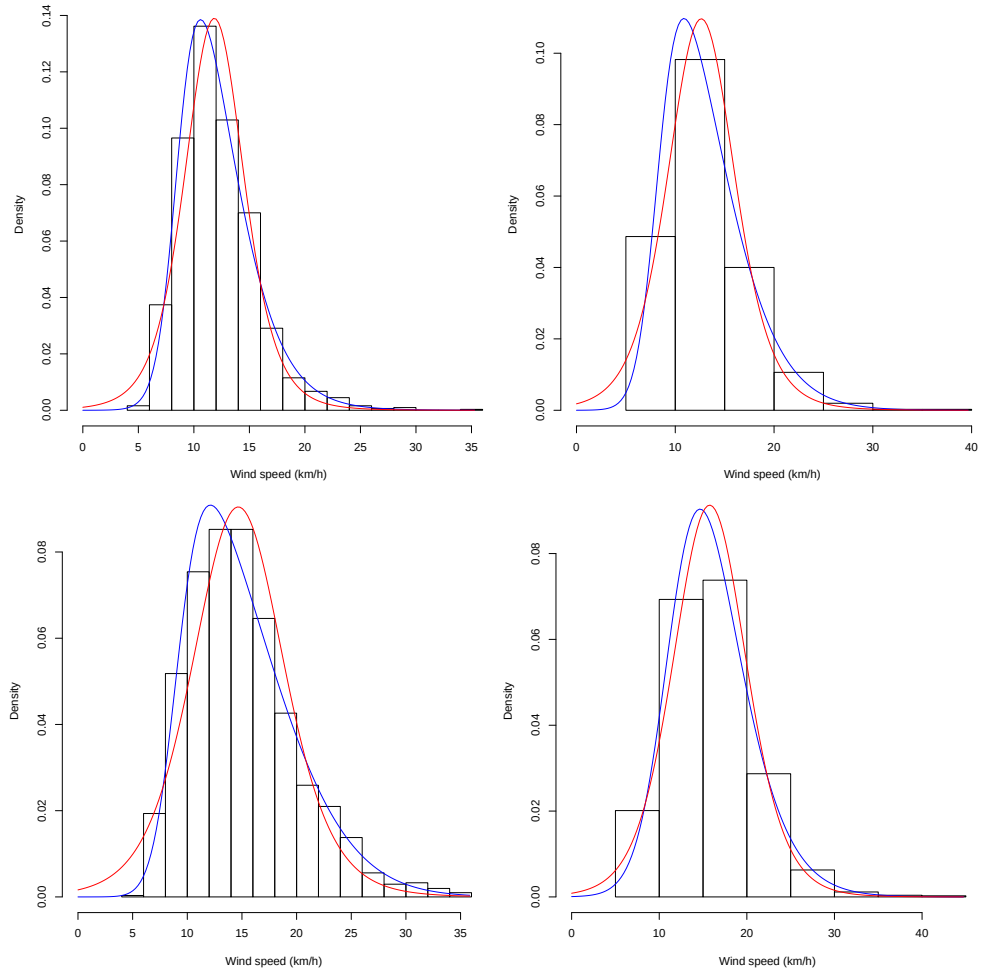

Figure 4.8: Histogram of the daily data, fitted t (red solid line) and skew-t (blue solid line) densities for the 4 stations with the largest slant coefficient.

Stations	Model	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\nu}$	CLIC
(CLOU,CLAY,SALL)	ex-t	$(0.67, 0.57, 0.69)$	–	2.89	5395.73
	ex-skew-t	$(0.42, 0.74, 0.52)$	$(-0.80, 2.88, -0.23)$	2.06	5385.07
(CLOU,CLAY,PAUL)	ex-t	$(0.59, 0.50, 0.69)$	–	2.53	5503.54
	ex-skew-t	$(0.45, 0.29, 0.65)$	$(-0.68, 21.07, 23.41)$	2.16	5496.90
(CLAY,SALL,PAUL)	ex-t	$(0.65, 0.61, 0.53)$	–	1.55	5086.13
	ex-skew-t	$(0.56, 0.51, 0.39)$	$(3.55, 2.36, 8.49)$	1.29	5075.87
(CLOU,SALL,PAUL)	ex-t	$(0.37, 0.40, 0.42)$	–	1.88	5428.04
	ex-skew-t	$(0.29, 0.30, 0.37)$	$(-0.14, 1.04, 34.70)$	2.11	5419.27

Table 4.2: Pairwise composite likelihood estimates $\hat{\vartheta} = (\hat{\omega}, \hat{\nu})$ and $\hat{\vartheta} = (\hat{\omega}, \hat{\alpha}, \hat{\nu})$ of the extremal-t (ext-t) and extremal skew-t (ex-skew-t) models respectively, for all possible triplets of the four locations CLOU, CLAY, PAUL and SALL.

argument of a better fit from the extremal skew-t model

For each location triple (X, Y, Z) we can also evaluate the conditional probability of exceeding some fixed threshold (x, y, z) using each parametric model. Table 4.3 presents estimated

probabilities of the two cases $\Pr(X > x|Y > y, Z > z)$ and $\Pr(X > x, Y > y|Z > z)$, along with the associated empirical probabilities and their 95% confidence intervals (CI) for a range of thresholds. For these specific thresholds, the extremal skew-t model provides estimates of the conditional probabilities that fall within the 95% empirical CI. However, four probabilities estimated with the extremal-t model are not consistent with the empirical CI. This indicates that the additional flexibility of the extremal skew-t model allows it to more accurately characterise the dependence structure evident in the observed data.

	Threshold	Extremal-t	Extremal skew-t	Empirical (95% CI)
$X\|Y,Z$	$(q^{90}_{\text{CO}}, q^{70}_{\text{CA}}, q^{70}_{\text{PA}})$	0.2587	0.2737	$0.3333\,(0.2706, 0.3960)$
	$(q^{90}_{\text{SA}}, q^{70}_{\text{CA}}, q^{70}_{\text{PA}})$	0.3268	0.3305	$0.2973\,(0.2356, 0.3590)$
	$(q^{90}_{\text{PA}}, q^{70}_{\text{CA}}, q^{70}_{\text{SA}})$	0.3752	0.3356	$0.2857\,(0.2247, 0.3467)$
	$(q^{90}_{\text{CO}}, q^{70}_{\text{SA}}, q^{70}_{\text{PA}})$	0.2686	0.3150	$0.3333\,(0.2706, 0.3960)$
$X,Y\|Z$	$(q^{90}_{\text{CO}}, q^{90}_{\text{CA}}, q^{70}_{\text{SA}})$	0.1196	0.0789	$0.0781\,(0.0420, 0.1142)$
	$(q^{90}_{\text{CA}}, q^{90}_{\text{PA}}, q^{70}_{\text{CO}})$	0.1236	0.0776	$0.0938\,(0.0546, 0.1330)$
	$(q^{90}_{\text{CO}}, q^{90}_{\text{SA}}, q^{70}_{\text{PA}})$	0.0896	0.1048	$0.0938\,(0.0550, 0.1326)$
	$(q^{90}_{\text{SA}}, q^{90}_{\text{PA}}, q^{70}_{\text{CO}})$	0.1038	0.1071	$0.0769\,(0.0415, 0.1123)$

Table 4.3: Extremal-t and extremal skew-t conditional probabilities of exceeding particular fixed thresholds of the form $\Pr(X > x|Y > y, Z > z)$ and $\Pr(X > x, Y > y|Z > z)$, along with empirical estimates. The windspeed thresholds (x, y, z) are constructed from the marginal quantiles $q^{70} = (q^{70}_{\text{CO}}, q^{70}_{\text{CA}}, q^{70}_{\text{SA}}, q^{70}_{\text{PA}}) = (18.04, 20.33, 24.18, 23.61)$ and $q^{90} = (q^{90}_{\text{CO}}, q^{90}_{\text{CA}}, q^{90}_{\text{SA}}, q^{90}_{\text{PA}}) = (22.11, 24.33, 29.05, 28.26)$ at each location.

Finally, Figure 4.9 provides examples of univariate (top panels) and bivariate (bottom) conditional return levels for each triple of sites. The return levels are computed conditionally on the wind at the remaining station(s) being higher than their upper 70% marginal quantile. The univariate conditional return levels of both models follow the observed data fairly well, and are mostly within the 95% empirical CI. Differences are that the extremal skew-t model return levels stay within the empirical CI more often in the rightmost panel, and that they more closely follow the most extreme observation in the centre panels compared to the extremal-t model.

The primary differences in the bivariate conditional return levels (bottom panels, Figure 4.9) are asymmetric contour levels obtained with the extremal skew-t model (blue line) in contrast with those symmetric obtained with the extremal-t model (red line). The difference is more noticeable in the leftmost and rightmost panel. As expected, the leftmost panel shows smaller return levels for the extremal skew-t model, and this is because (CLOU, SALL) have negative slant parameters and so the joint tail is shorter than that of the extremal-t. Conversely, the rightmost panel exhibits larger return levels for the extremal skew-t model, and this is because (CLOU, PAUL) have a slant parameter that is close to zero and a very large slant parameter and so the joint tail is longer than that of the extremal-t. The slight differences in the middle panels may be explained as follows. The slant parameters of (CLAY, PAUL,SALL) are small positive values and hence there is not much difference between the joint tails of extremal skew-t and t models. The slant parameters of (CLOU,PAUL) are a large positive value and one slightly negative. However, because the parameter of CLAY is also a large positive value this also makes

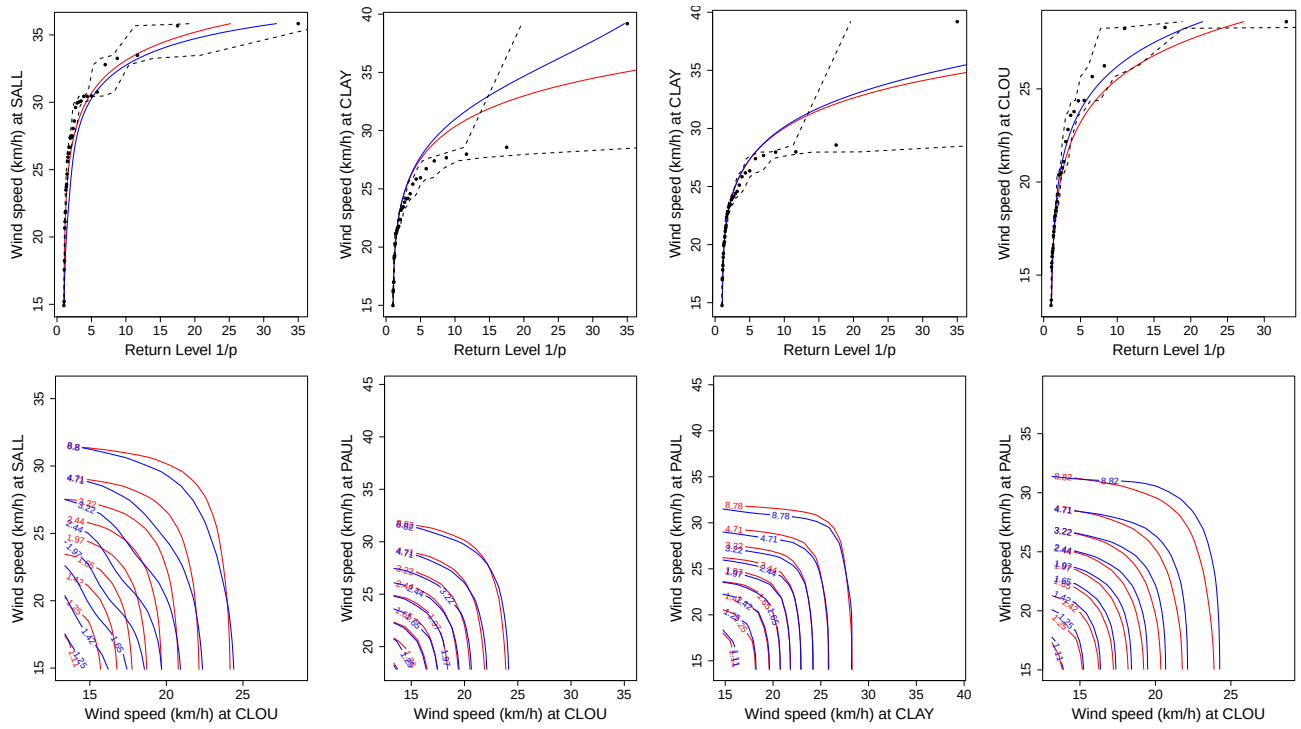

Figure 4.9: Univariate (top row) and bivariate (bottom) conditional return levels for the triples (left-to-right): (CLOU, CLAY, SALL), (CLOU, CLAY, PAUL), (CLAY, SALL, PAUL) and (CLOU, SALL, PAUL). Red and blue lines respectively indicate return levels calculated from extremal-t and extremal skew-t models. Points indicate the empirical observations and the black dashed lines their 95% confidence interval.

that there is not much difference between the joint tails of the two models.

In summary, for these wind speed data, the more flexible extremal skew-t model is demonstrably superior to the extremal-t model in describing the extremes of both the univariate marginal distributions, and the extremal dependence between locations.

4.6 Discussion

Appropriate modelling of extremal dependence is critical for producing realistic and precise estimates of future extreme events. In practice this is a hugely challenging task, as extremes in different application areas may exhibit different types of dependence structures, asymptotic dependence levels, exchangeability, and stationary or non-stationary behaviour.

Working with families of skew-normal distributions and processes we have derived flexible new classes of extremal dependence models. Their flexibility arises as they include a wide range of dependence structures, while also incorporating several previously developed and popular models, such as the stationary extremal-t process and its sub-processes, as special cases. These include dependence structures that are asymptotically independent, which is useful for describing the dependence of variables that are not exchangeable, and a wide class of non-stationary, asymptotically dependent models, suitable for the modelling of spatial extremes.

In terms of future development, semi-parametric estimation methods would provide powerful

techniques to fully take advantage of the flexibility offered by non-stationary max-stable models. Such methods can be computationally demanding, however. An interesting further direction would be to design simple and interpretable families of covariance functions for skew-normal processes for which it is then possible to derive max-stable dependence models that are useful in practical applications.

Chapter 5

Exploratory data analysis for extreme values using non-parametric kernel methods

Abstract

Non-parametric kernel methods are applied to the analysis of univariate and multivariate extreme values data. As kernel estimators possess excellent visualisation properties, they allow for more informative visualisations beyond those provided by scatter plots or histograms for graphical exploratory data analysis. As further analysis of extreme values data relies heavily on parametric models, considerable efforts have expended in developing and understanding a wide range of these models. Kernel estimators, as they place minimal assumptions on the data, are well-placed to measure the appropriateness of a purported parametric model to the data sample, and to improve model selection over currently available histogram based comparisons. We verify the performance of these kernel based analyses for simulated and for experimental climate data. For the latter, we investigate the ability of well-known geophysical climate models to predict the observed extreme temperatures in Sydney, Australia during the twentieth century.

5.1 Introduction

Extreme values (very large or very small values) are of particular interest as they are closely related to highly unlikely events with important consequences. For climate data, these extreme events include heat waves (extreme high temperatures), cold snaps (extreme low temperatures), floods (extreme high levels of waterways or tides or waves), storms (extreme high wind speeds or amounts of precipitation), droughts (extreme low amounts of rainfall). For extreme values in other contexts, see Kotz and Nadarajah (2000); Coles (2001). Historically univariate quantities have been the most extensively studied. Let X be a random variable with cumulative distribution function (c.d.f.) F_X with probability density function (p.d.f.) f_X. We focus on values which are greater than a threshold u, denoted as $X^{[u]} \equiv X|X > u$, with the corresponding tail distribution

$F_{X^{[u]}}$ and tail density $f_{X^{[u]}}$. The threshold u, which is unknown but non-random, determines the support on which values of X are considered to be extreme values.

Our goal is to characterise the tail behaviour of X by estimating the tail density $f_{X^{[u]}}$. Common methods rely on parametric models based on the Generalised Extreme Value (GEV) distributions. Numerous estimation procedures are available for GEV distributions. The most accurate methods are maximum likelihood, studied by Prescott and Walden (1980); Hosking (1985); Smith (1985) and Macleod (1989), and the probability weighted moments as considered by Hosking et al. (1985). Cheng and Amin (1983) proposed an alternative using maximum product spacing. Other methods include least squares estimation (Maritz and Munro, 1967), estimation based on order statistics and records (Pickands, 1975; Hill, 1975), method of moments (Christopeit, 1994), and Bayes estimation (Lye et al., 1993). As with all parametric estimators, these suffer from the possibility of misspecification. In order to avoid potential misspecification, we rely on non-parametric estimation which does not require a pre-specified parametric form. Markovich (2007, Chapter 3) provides a summary of non-parametric estimation of univariate heavy-tail densities. We use kernel estimators, as they are amongst the most widely used non-parametric estimators, see Wand and Jones (1995); Simonoff (1996) for an overview. Kernel density estimators possess excellent visualisation properties which we incorporate within a graphical exploratory analysis of extreme values data. However they can produce spurious bumps in the tails of f_X if it has heavy tails as is the case for extreme value analysis. Our proposed approach focuses on modifications which attenuate these spurious bumps to obtain an estimate of the tail density $f_{X^{[u]}}$. Our first proposition is to focus on the truncated sample greater than a threshold which resolves (or at least reduces) the appearance of these spurious bumps. A second proposition is to apply transformation kernel estimators for tail densities. Transformation kernel estimators based on log transformations are well-known already as they map a bounded support to the unbounded Euclidean space, allowing the use of standard kernel estimation in the transformed space. The crucial question of smoothing parameter selector has not been addressed explicitly yet as it appears to be implied that smoothing parameter selection and estimation are performed in the unbounded space and then back-transformed to the original space. We establish the mean integrated squared error optimality of this implied approach.

A further advantage of kernel estimators is their straightforward extension to multivariate data analysis. For multivariate extremes, non-parametric estimation of various indicators of extremal dependence has been and is still currently an intensively studied field. The Pickands or extremal dependence function (see Pickands (1975); Hall and Tajvidi (2000); Marcon et al. (2014) among others), the tail dependence function (Huang, 1992; Drees and Huang, 1998; Einmahl et al., 2008, 2012), the spectral measure (Einmahl et al., 2001; Einmahl and Segers, 2009; de Carvalho et al., 2013) are some examples of a growing interest. In contrast to these authors who concentrate on quantitative measures, our primary focus is to provide exploratory data analysis tools. This multivariate framework allows us to examine phenomena which are simultaneously extreme in several different variables. As an example, many climate models forecast a rise in the mean temperature for many locations on the Earth's surface. Upon more detailed inspection, some of them predict that this overall increase in the mean temperature will

be accompanied by a simultaneous increase in warmer minimum and maximum temperatures. Furthermore, as there are many competing climate climate models, a model selection procedure that determines which of these models is most consistent with the empirical data is an important tool. In this article we define a divergence or goodness-of-fit measure on the integrated difference of the tail estimator and a hypothesised density, which is then used subsequently for model selection. It is a generalisation of the goodness-of-fit measure in Perkins et al. (2007) which compared an estimated and a hypothesised histogram for model selection. Our definition is sufficiently general in that it is not tailored only to kernel estimators but can be applied to any integrable density estimator (including histograms).

The layout of this article is as follows: we introduce tail samples for univariate and multivariate data in Section 5.2, as well histogram density estimators and our proposed transformation kernel density estimator. We also analyse the asymptotic behaviour of these two broad classes of density estimators when using data-based methods to select the optimal amount of smoothing for tail density estimation. We then proceed to develop a model selection procedure based on the discrepancy between non-parametric density estimators to a suite of parametric models. In Section 5.3 we verify our results on finite samples for simulated and experimental data and Section 4 concludes with a discussion. All mathematical proofs are deferred to the Appendix.

5.2 Tail densities for extreme values

5.2.1 Density estimation

Let $X_1, \ldots, X_m$ be a random sample drawn from the common univariate distribution F_X with density f_X. Let n of these observations exceed the threshold u, to form the tail sample $X_1^{[u]}, \ldots, X_n^{[u]}$ where $X_i^{[u]} \in (u, \infty), i = 1, \ldots, n$. Standard kernel estimators are most often defined for data with unbounded support. Using these kernel estimators without modification on bounded data like $\mathcal{X}^{[u]}$ can lead to increased bias at the boundary of the support. One major class of modified estimators are based on the modifying the kernel function itself to avoid assigning probability mass outside the data support, e.g. the linear boundary kernel Gasser and Müller (1979) or the boundary beta kernel Chen (1999). We do not pursue these modifications and instead we focus on an alternative modification based on transformation kernel estimators where a known monotonic transformation t maps the data support to the real line where standard kernel estimators are well-established, before back-transforming to the original bounded support, see Devroye and Györfi (1985); Silverman (1986); Wand and Jones (1995).

Let $Y = t(X^{[u]})$ be a transformed random variable, with distribution F_Y and density f_Y. The relationship between the transformed random variable Y and the original $X^{[u]}$ at a non-random point x is given by

$$f_{X^{[u]}}(x) = |t'(x)| f_Y(t(x))$$

where t' is the first derivative of t. Instead of directly approaching the difficult task of estimating $f_{X^{[u]}}$, we approach it indirectly via the simpler task of estimating f_Y. Consider the transformed sample $Y_1, \ldots, Y_n$ where $Y_i = t(X_i), i = 1, \ldots, n$, and $y = t(x)$. Since $X_1^{[u]}, \ldots X_n^{[u]}$ are supported

on (u, ∞), a suitable transformation would be $t(x) = \log(x - u)$. For data-based choices of the transformation function, see Ruppert and Wand (1992); Ruppert and Cline (1994). We do not need to pursue these as the logarithm function is adequate for our purposes. Now that $Y_1, \ldots, Y_n$ are supported on the real line, f_Y can be estimated by the standard kernel density estimator

$$\hat{f}_Y(y; h) = n^{-1} \sum_{i=1}^{n} K_h(y - Y_i)$$

where $K_h(y) = h^{-1}K(y/h)$ is a scaled kernel, $h > 0$ is the bandwidth or smoothing parameter and K is a symmetric kernel density function. The tail density estimator can be defined by replacing the true density by its kernel estimator

$$\hat{f}_{X^{[u]}}(x; h) = |t'(t^{-1}(y))| \hat{f}_Y(y; h).$$

A generalisation of this transformation kernel estimator to multivariate data is as follows. Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-dimensional random variable with distribution $F_{\boldsymbol{X}}$ with density $f_{\boldsymbol{X}}$. The random variable of values greater than a vector threshold $\boldsymbol{u} = (u_1, \ldots, u_d)$ is denoted as $\boldsymbol{X}^{[u]} \equiv \boldsymbol{X} | \boldsymbol{X} > \boldsymbol{u}$ where we mean that each marginal inequality must hold, i.e. $X_j > u_j$ for $j = 1, \ldots, d$. The support of $\boldsymbol{X}^{[u]}$ is the Cartesian product $(\boldsymbol{u}, \infty) = (u_1, \infty) \times \cdots \times (u_d, \infty)$. For $\boldsymbol{x} \in (\boldsymbol{u}, \infty)$, the corresponding tail density is $f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}) / \bar{F}_{\boldsymbol{X}}(\boldsymbol{u})$ and tail distribution is $F_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(\boldsymbol{x}) / \bar{F}_{\boldsymbol{X}}(\boldsymbol{u})$, where $\bar{F}_{\boldsymbol{X}}(\boldsymbol{u}) = \int_{(\boldsymbol{u}, \infty)} f_{\boldsymbol{X}}(\boldsymbol{w}) d\boldsymbol{w}$ is the survival function of $\boldsymbol{X}$ evaluated at $\boldsymbol{u}$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be a random sample drawn from the common d-variate distribution $F_{\boldsymbol{X}}$ with density $f_{\boldsymbol{X}}$. Let n of these observations exceed the threshold $\boldsymbol{u}$, to form the tail sample $\boldsymbol{X}_1^{[\boldsymbol{u}]}, \ldots, \boldsymbol{X}_n^{[\boldsymbol{u}]}$. Consider the transformed random variable $\boldsymbol{Y} = \boldsymbol{t}(\boldsymbol{X}^{[u]})$ where $\boldsymbol{t} : (\boldsymbol{u}, \infty) \to \mathbb{R}^d$ is defined by $\boldsymbol{t}(\boldsymbol{x}) = (t_1(x_1), \ldots, t_d(x_d))$ where the t_j are monotonic functions on (u_j, ∞), e.g. $t_j(x_j) = \log(x_j - u_j), j = 1, \ldots, d$. The density of $\boldsymbol{X}^{[u]}$ is related to the density of $\boldsymbol{Y}$ by

$$f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) = f_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x})) |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|$$

where $|\mathbf{J}_{\boldsymbol{t}}|$ is Jacobian of $\boldsymbol{t}$. The transformed data sample are $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, with $\boldsymbol{Y}_i = \boldsymbol{t}(\boldsymbol{X}_i), i = 1, \ldots, n$. The kernel estimator of $f_{\boldsymbol{Y}}$ at a non-random point $\boldsymbol{y} = (y_1, \ldots, y_d) = \boldsymbol{t}(\boldsymbol{x})$ is given by

$$\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}; \mathbf{H}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\boldsymbol{y} - \boldsymbol{Y}_i)$$

where K is a symmetric d-variate density function, $\mathbf{H}$ is the bandwidth matrix which is a $d \times d$ positive definite symmetric matrix of smoothing parameters, and the scaled kernel $K_{\mathbf{H}}(\boldsymbol{y}) = |\mathbf{H}^{-1/2}| K(\mathbf{H}^{-1/2} \boldsymbol{y})$. Note that the bandwidth matrix corresponds to the squared data scale, whereas the scalar bandwidth corresponds to the data scale, e.g. for the Gaussian kernel, $\mathbf{H}$ is the variance matrix, and h is the standard deviation. The tail density can be defined by replacing the true density function by its kernel estimator

$$\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}; \mathbf{H}) = |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{t}^{-1}(\boldsymbol{y}))| \hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}; \mathbf{H}) \tag{5.1}$$

where $\boldsymbol{t}^{-1}(\boldsymbol{y}) = (t_1^{-1}(y_1), \ldots, t_d^{-1}(y_d))$ is the element-wise inverse of $\boldsymbol{t}(\boldsymbol{y})$. Under the regularity conditions (A1)–(A4) in the Appendix, the mean integrated squared error (MISE) of the density estimator $\hat{f}_{\boldsymbol{Y}}$ is

$$\operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) = \mathbb{E} \int_{\mathbb{R}^d} [\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}; \mathbf{H}) - f(\boldsymbol{y})]^2 \mathrm{d}\boldsymbol{y} = O(\operatorname{tr} \mathbf{H}^2 + n^{-1}|\mathbf{H}|^{-1/2})$$

using established procedures (see Lemma E.1 in the Appendix). That the MISE of $\hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ is the same order follows from immediately from Equation (5.1). Using these MISE expressions, we can define optimal bandwidth selectors as the minimiser of the MISE,

$$\mathbf{H}^* = \operatorname*{argmin}_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) = O(n^{-2/(d+4)}) \tag{5.2}$$

where $\mathcal{F}$ is the space of $d \times d$ symmetric positive definite matrices.

The following theorem asserts that we can carry out all the important steps of bandwidth selection and estimation for the transformed data with unbounded support, allowing us to take advantage of existing results and algorithms, and that this transformation does not affect the asymptotic behaviour, and so we are able to recover the optimal convergence rates of the unbounded data estimator.

Theorem 5.1. *Suppose that the conditions (A1–A4) hold. The minimal MISE rate of the tail density estimator $f_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ is asymptotically the same order as the minimal MISE rate of the transformed kernel density estimator $\hat{f}_{\boldsymbol{Y}}$, as $n \to \infty$,*

$$\inf_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H}) - \Big\{ \inf_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) \Big\} = O(n^{-4/(d+4)}).$$

On the other hand, if we compute a standard kernel estimator of type $\hat{f}_{\boldsymbol{Y}}$ to the untransformed bounded tail sample $\boldsymbol{X}_1^{[\boldsymbol{u}]}, \ldots, \boldsymbol{X}_n^{[\boldsymbol{u}]}$, the MISE is $O(n^{-1}|\mathbf{H}|^{-1/2} + \operatorname{tr} \mathbf{H})$ which leads to a minimal MISE rate of $O(n^{-2/(d+2)})$ which is asymptotically slower than the minimal MISE rates of $O(n^{-2/(d+4)})$ in Theorem 5.1, see Marshall and Hazelton (2010). This implies that it is important to take the boundedness of the tail sample into account.

The optimal bandwidth selector defined in Equation (5.2) is mathematically intractable as it depends on unknown quantities. A vast body of research has thus been constructed to provide data-based optimal bandwidth selectors which estimate/approximate these quantities. There are three main classes: (i) normal scale (or rule of thumb), (ii) plug-in and (iii) cross validation.

The class of normal scale selectors is an extension to the multivariate case of the quick and simple bandwidth selectors where the unknown density f is replaced by a normal density,

$$\hat{\mathbf{H}}_{\mathrm{NS}} = \left[\frac{4}{(d+2)n} \right]^{2/(d+4)} \mathbf{S} n^{-2/(d+4)},$$

where $\mathbf{S}$ is the sample variance matrix of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, see Wand and Jones (1995, p. 111).

The class of plug-in selectors consists of a generalisation of the work of Sheather and Jones (1991) by Wand and Jones (1994); Duong and Hazelton (2003). Plug-in selectors use as a starting

point the AMISE formula where the only unknown quantity is the $\boldsymbol{\psi}_{\boldsymbol{Y},4} = \int_{\mathbb{R}^d} \mathsf{D}^{\otimes 4} f_{\boldsymbol{Y}}(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$ functional. The fourth order differential $\mathsf{D}^{\otimes 4}$ is expressed as a vector of length d^4, resulting from four-fold Kronecker product of the first order differential D, see Holmquist (1996). Replacing this by an estimator $\hat{\boldsymbol{\psi}}_{\boldsymbol{Y},4}$ yields a plug-in criterion

$$\mathrm{PI}(\mathbf{H}) = \tfrac{1}{4} m_2^2(K) (\mathrm{vec}^\top \mathbf{H} \otimes \mathrm{vec}^\top \mathbf{H}) \hat{\boldsymbol{\psi}}_{\boldsymbol{Y},4}(\mathbf{G}) + n^{-1} R(K) |\mathbf{H}|^{-1/2}$$

where $m_2(K)$ is defined in condition (A2), vec is the operator that stacks the element of a matrix column-wise into a vector, $\hat{\boldsymbol{\psi}}_{\boldsymbol{Y},4}(\mathbf{G}) = n^{-2} \sum_{i,j=1}^n \mathsf{D}^{\otimes 4} L_{\mathbf{G}}(\boldsymbol{Y}_i - \boldsymbol{Y}_j)$, $L_{\mathbf{G}}$ being the pilot kernel with pilot bandwidth matrix $\mathbf{G}$ and $R(K) = \int_{\mathbb{R}^d} K(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x}$. The plug-in selector $\hat{\mathbf{H}}_{\mathrm{PI}}$ is the minimiser over $\mathcal{F}$ of $\mathrm{PI}(\mathbf{H})$.

For the class of cross validation selectors we focus on unbiased (or least squares) cross validation and smoothed cross validation. Unbiased cross validation (UCV) was introduced by Bowman et al. (1984) and Rudemo (1982). The unbiased cross validation selector, $\hat{\mathbf{H}}_{\mathrm{UCV}}$, is defined as the minimiser over $\mathcal{F}$ of

$$\mathrm{UCV}(\mathbf{H}) = \int_{\mathbb{R}^d} \hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}; \mathbf{H})^2 \mathrm{d}\boldsymbol{x} - 2n^{-1} \sum_{i=1}^n \hat{f}_{\boldsymbol{Y},-i}(\boldsymbol{Y}_i; \mathbf{H}),$$

where $\hat{f}_{\boldsymbol{Y},-i}(\boldsymbol{Y}_i; \mathbf{H}) = [n(n-1)]^{-1} \sum_{j=1}^n K_{\mathbf{H}}(\boldsymbol{Y}_i - \boldsymbol{Y}_j)$. The smoothed cross validation (SCV) selector $\hat{\mathbf{H}}_{\mathrm{SCV}}$, is defined as the minimiser over $\mathcal{F}$ of

$$\begin{aligned}\mathrm{SCV}(\mathbf{H}) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n & (K_{\mathbf{H}} * K_{\mathbf{H}} * L_{\mathbf{G}} * L_{\mathbf{G}} - 2K_{\mathbf{H}} * L_{\mathbf{G}} * L_{\mathbf{G}} + L_{\mathbf{G}} * L_{\mathbf{G}})(\boldsymbol{Y}_i - \boldsymbol{Y}_j) \\ & + n^{-1} R(K) |\mathbf{H}|^{-1/2},\end{aligned}$$

where $*$ is the convolution operator, see Hall et al. (1992). If there are no replications in the data, then SCV with $\mathbf{G} = 0$ is identical to UCV as L_0 can be thought of as the Dirac delta function.

The UCV selector can be computed numerically as it contains no unknown quantities. On the other hand, the question of the pilot bandwidth remains for the plug-in and SCV selectors. For computational data-based algorithms which resolve this question, see Wand and Jones (1995); Duong and Hazelton (2003) for plug-in selectors and Hall et al. (1992); Duong and Hazelton (2005) for SCV selectors.

We conclude this section by defining a histogram. Histograms, especially for univariate data, are widely used as alternatives to kernel estimators for visualising data samples. Their advantages include a simple and computationally easy method of computation and that they do not suffer from the boundary bias problems of standard kernel estimators. For the tail sample $\boldsymbol{X}_1^{[\boldsymbol{u}]}, \ldots, \boldsymbol{X}_n^{[\boldsymbol{u}]}$, divide the data range into a regular partition of hypercubes A_i of size $b_1 \times \cdots \times b_d$. Let $\boldsymbol{b} = (b_1, \ldots, b_d) \in \mathbb{R}^d$ be the binwidth. The histogram estimator of $\tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ at a point $\boldsymbol{x}$ in a bin A_i is

$$\tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\boldsymbol{x}; \boldsymbol{b}) = \frac{\gamma_i}{n b_1 \cdots b_d}$$

where γ_i represents the number of observations in the hypercube A_i. If conditions similar to (A1) and (A3) are fulfilled then, by Scott (2015, Theorem 3.5), the MISE of the histogram estimator is

$$\text{MISE}\, \tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \boldsymbol{b}) = O((nb_1 \cdots b_d)^{-1} + (b_1^2 + \cdots b_d^2))$$

with minimal MISE

$$\inf_{\boldsymbol{b}>0} \text{MISE}\, \tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \boldsymbol{b}) = O(n^{-2/(d+2)}).$$

which is asymptotically slower than $O(n^{-4/(d+4)})$ minimal MISE rate for the transformation kernel estimator $\tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ from Theorem 5.1. So from a MISE point of view, the transformation kernel estimator is preferred to a histogram, especially as the dimension d increases.

Analogously to the data-based optimal bandwidth selectors, the normal scale optimal binwidth, from Scott (2015, Theorem 3.5), is

$$\hat{b}_j^* = 2 \times 3^{1/(d+2)} \pi^{d/(d+4)} s_j n^{-1/(d+2)} \tag{5.3}$$

where $s_j, j = 1, \ldots, d$ are the marginal sample standard deviations of $\boldsymbol{X}_1^{[\boldsymbol{u}]}, \ldots, \boldsymbol{X}_n^{[\boldsymbol{u}]}$. There have not been the equivalent variety of binwidth selectors which generalise Equation (5.3) compared to bandwidth selectors due the slower asymptotic performance of histograms as compared to kernel estimators.

5.2.2 Model selection

Suppose that we have a suite of M parametric models, and we wish to determine which of them is the most appropriate fit to an extreme values data set. Since the target underlying model is unknown, we do not have a well-defined target to which to compare these parametric models. Perkins et al. (2013) introduce the histogram estimator $\tilde{f}_{\boldsymbol{X}^{[u]}}$ of the observed data sample as the surrogate for the unknown target, and so the fit of the parametric models is assessed according to the discrepancy of the induced parametric density functions $g_1, \ldots, g_M$ and the histogram $\tilde{f}_{\boldsymbol{X}^{[u]}}$. Their tail index is given by

$$\tilde{T}(g_j) = \int_{(\boldsymbol{u},\infty)} |g_j(\boldsymbol{x}) - \tilde{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\boldsymbol{x})| \mathrm{d}\boldsymbol{x} \tag{5.4}$$

where $(\boldsymbol{u}, \infty)$ is short hand for the Cartesian product $(u_1, \infty) \times \cdots \times (u_d, \infty)$. The selected model is the one which gives the smallest discrepancy

$$\operatorname*{argmin}_{j \in \{1,\ldots,M\}} \tilde{T}(g_j).$$

Our proposal is to replace the histogram in Equation (5.4) with a kernel estimator. The artefacts induced by a histogram grid are avoided or least reduced, namely the anchor point problem to choose where the histogram grid begins and the empty bin problem where it is not clear whether histogram bins with empty counts should be interpreted as a true zero probability or are due to insufficient observed data. This latter is important for extreme values as they are

sparsely distributed in the tail regions. We can proceed to calculate the tail indices $\hat{T}$ as

$$\hat{T}(g_j) = \int_{(\boldsymbol{u},\infty)} |g_j(\boldsymbol{x}) - \hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x})| \mathrm{d}\boldsymbol{x}, \tag{5.5}$$

Additionally, in order to highlight the purpose of working with transformed histogram, the following standard kernel is considered for comparison:

$$\hat{T}^*(g_j) = \int_{(\boldsymbol{u},\infty)} |g_j(\boldsymbol{x}) - \hat{f}_{\boldsymbol{X}}(\boldsymbol{x})| \mathrm{d}\boldsymbol{x}, \tag{5.6}$$

where $\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})$ represents the standard kernel density estimator. Again the selected models by both kernel based tail indices are the ones which give the smallest discrepancy. The integrals in Equations (5.4)–(5.6) are usually approximated by (weighted) Reimann sums.

5.3 Numerical results

5.3.1 Simulated data - univariate

A first set of simulations is carried out in order to show that the kernel density estimator introduced in Section 5.2.1 is a good surrogate for the target density of extreme values. Secondly, we carry out the model selection procedure in Section 5.2.2. These two simulation studies will be carried out considering both univariate and bivariate data.

For the target densities, we use the well known 3-parameter Generalized Extreme Value (GEV) and Generalized Pareto (GPD) distributions to generate simulated data. For the former, the critical parameter is the third parameter which distinguishes the GEV distribution into the Gumbel, Fréchet and Weibull distributions. For each target density, we can compute (at least numerically) the target tail density $f_{X^{[u]}}(x) = f_X(x)/\bar{F}_X(u)$. We generate a sample of size $m = 2000$ and set the threshold u at the 95% upper quantile to compute the tail sample of size $n = 100$. For each sample, we compute the:

1. transformation kernel based estimators with the plug-in (KPI), unbiased cross validation (KUC), smoothed cross validation (KSC), and normal scale (KNS) selectors, all with the transformation $t(x) = \log(x - u)$
2. standard kernel based estimators with the plug-in (KPI*), unbiased cross validation (KUC*), smoothed cross validation (KSC*), and normal scale (KNS*) selectors.
3. histogram estimator with normal scale binwidth (HIS)
4. parametric estimators: Fréchet (FRE), Gumbel (GUM), Generalized Pareto (GPD).

The top row of Figure 5.1 provides an example of tail density estimation from three different distributions, when $n = 100$. From left to right these distributions are Fréchet (with location 1, scale 0.5 and shape 0.25), Gumbel (with location 1.5 and scale 3) and GPD (with location 0, scale 1 and shape 0.25). The target density function is the solid black curve, the histogram is the dashed red, and the plug-in kernel estimators are the dotted green (transformed) and the dot-dashed blue (standard), where we display only the plug-in estimators for clarity. The

transformed kernel estimators are more accurate estimators of the tail behaviour, exhibiting less bias at the boundary of the support than standard kernel estimators. Also they have the advantages of being continuous kernel estimators leading to better visualisation and are more helpful when comparing it to a continuous target density. The optimal histogram binwidth tends to be too small for these tail samples and so produces histograms which are more noisy than the kernel estimators. The bottom row of Figure 5.1 provides diagnostics for extremes through qq-plots of the target quantiles versus the non-parametric estimated quantiles. The histogram HIS quantiles most consistently follow the 45° line, though the transformed kernel KPI quantiles are more consistent for the target Gumbel. For the other two target densities, the KPI quantiles are uniformly larger than the HIS quantiles. The standard kernel KPI* quantiles either under- or over-estimate the target quantiles. As expected, histogram quantiles are more accurate than the quantiles of the kernel density estimators, as the the latter aim to optimally estimate the density function rather than the quantile function. Despite this under-performance of the KPI estimator, it is convenient to recall here that their smoothness provides a more informative graph, e.g. for visualising tangent slopes.

For more quantitative results, we generate 400 samples of size $m = 500, 1000$ and 2000 and set the threshold u at the 95% upper quantile to compute the tail sample of size $n = 25, 50, 100$. As these three sample sizes gave the similar results, we only present those for $n = 100$ for brevity. The measures of accuracy of an estimator will be the numerical approximation (Reimann sum) of the L_2 error $\int_{(\boldsymbol{u},\infty)}[\text{estimator of} f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) - f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x})]^2 \,\mathrm{d}\boldsymbol{x}$, and of the χ^2 error $\int_{(\boldsymbol{u},\infty)}(\text{estimator of} f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) - f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}))^2 / f_{\boldsymbol{X}^{[u]}}$. Other common diagnostics for extremes such as the Anderson-Darling divergence and extremal coefficients could also be computed but are omitted here as the estimation of the cumulative distribution function is required.

In Figure 5.2, box-plots show the accuracy of each estimator of the tail density depending on the model used to simulate the data, respectively Fréchet, Gumbel and Generalized Pareto, from left to right. The top row considers the L_2 error while the measure of error of the bottom row is calculated using the χ^2 error. Within each panel, the box-plots respectively correspond to the Fréchet (FRE), Gumbel (GUM), Generalized Pareto (GPD), Histogram (HIS), plug-in transformed kernel (KPI), unbiased cross validation transformed kernel (KUC), smoothed cross validation transformed kernel (KSC) and normal scale transformed kernel (KNS) estimators. The "$*$" symbol indicates the standard kernel estimator. As expected, for each distribution the most accurate density estimator is the correctly specified parametric model. Transformed kernel estimators systematically perform better than the histogram estimator even though it can produce larger variability (especially in the case of the GPD). Furthermore the standard kernel estimator is consistently producing the worst estimate of the tail density with however an improvement compared to other estimators for the GPD case. The differences in the accuracy between the kernel estimators with different bandwidth selectors is small, in contrast to other simulation studies where the bandwidth selection class is a crucial factor, see for example, Sheather and Jones (1991) and Wand and Jones (1995, Chapter 3), as these extreme values densities do not have complex, multimodal structures. Both measures of discrepancy provide similar results.

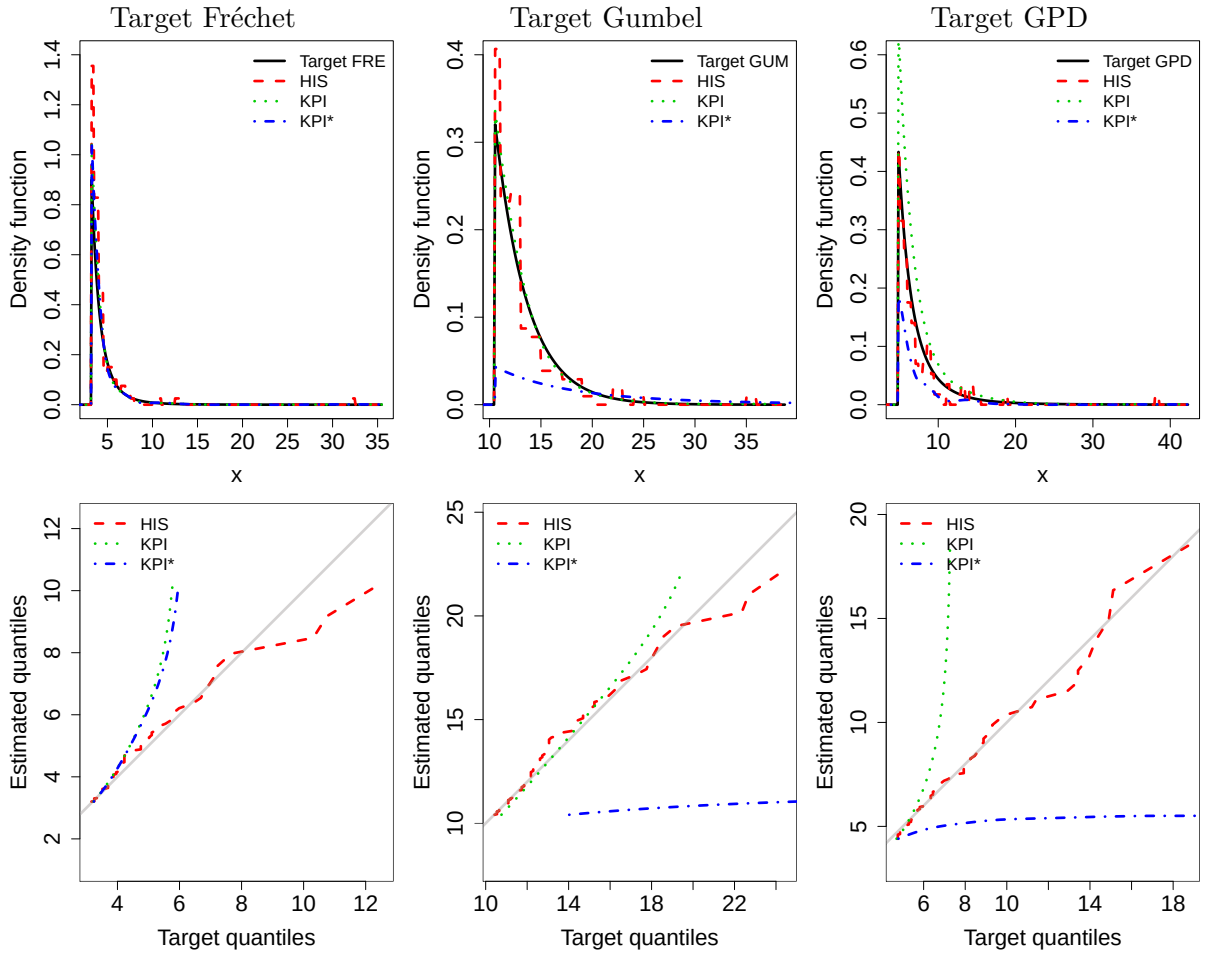

Figure 5.1: Non-parametric estimators of the univariate tail density (top) and of the tail quantiles (bottom) when the target density is Fréchet (left), Gumbel (middle) and GPD (right). Sample size is $n = 100$. Target density is represented by a solid black line. Histogram estimator with normal scale binwidth (HIS) is represented by a dashed red line, the transformed kernel plug-in estimator (KPI) by a doted green line and the standard kernel estimator (KPI*) by a dot-dashed blue line.

In Table 5.1, for each of the three target densities considered earlier and for sample size $n = 100$, we compute the tail indices $\tilde{T}$ (histogram), $\hat{T}$ (transformed kernel) and $\hat{T}^*$ (standard kernel) with respect to each parametric model. The tail indices are computed using both the L_2 and χ^2 errors (respectively top and bottom rows). The proportion of the 400 simulation trails where the target model is correctly selected is given in bold. Because the Fréchet and Gumbel distributions are nested we perform a deviance test. If the smaller model (Gumbel) provides a significantly better fit than the larger model (Fréchet), meaning that the shape parameter is significantly not different from zero, then we only consider the Gumbel and GPD fits. The model selection for the L_2 error is overall more accurate than the χ^2 error. For any of the three distributions, using kernel based estimators as a surrogate for the target density $\hat{T}$ selects the correct target in the vast majority of cases, with proportions higher than with a histogram tail index $\tilde{T}$, except for the standard kernel based tail index when the target density is GPD. With the L_2 error, when the target density is Fréchet, the histogram based tail index performs very

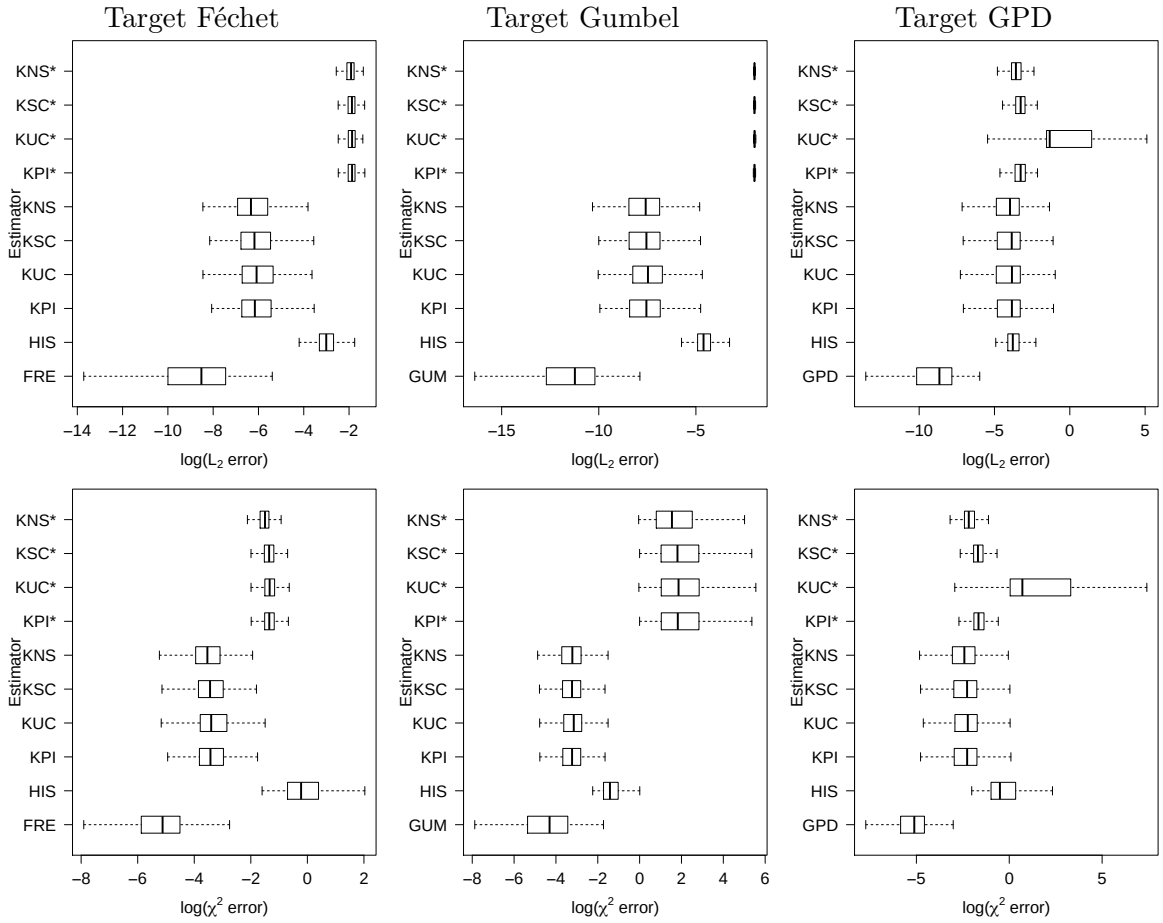

Figure 5.2: Box-plots of the $\log L_2$ (top) and $\log \chi^2$ (bottom) errors for the univariate Fréchet (FRE), Gumbel (GUM), Generalized Pareto (GPD), Histogram (HIS), plug-in kernel (KPI), unbiased cross validation kernel (KUC), smoothed cross validation kernel (KSC) and normal scale kernel (KNS) tail density estimators, for sample size $n = 100$. From left to right, the target densities are Fréchet, Gumbel and GPD.

poorly. The transformation kernel based index exhibits more stability than the standard kernel, consistently selecting the right model with high proportion. This inaccurate histogram model selection is corrected when the χ^2 error is considered, though at the expense of reducing the accuracy with the GPD target.

5.3.2 Simulated data - multivariate

The analysis of multivariate extremes is a challenging area where numerous approaches have been considered and applied to real examples. See for example Kotz and Nadarajah (2000); Coles (2001); Beirlant et al. (2004); de Haan and Ferreira (2006); Falk et al. (2011) for theoretical details and applications. The well known and commonly used GEV distribution available in the univariate case unfortunately no longer exist in higher dimensions. The study of multivariate extremes is then decomposed into two parts: first the margins are assumed to be GEV distributed and then transformed to unit Fréchet. This allows us to make the assumption that the remaining

		$\tilde{T}$ selected model			$\hat{T}$ selected model			$\hat{T}^*$ selected model		
Error	Target	FRE	GUM	GPD	FRE	GUM	GPD	FRE	GUM	GPD
L_2	FRE	**0.03**	0.18	0.79	**0.92**	0.00	0.08	**0.89**	0.00	0.11
	GUM	0.02	**0.90**	0.08	0.04	**0.94**	0.02	0.02	**0.98**	0.00
	GPD	0.01	0.28	**0.71**	0.11	0.05	**0.84**	0.00	0.31	**0.69**
χ^2	FRE	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
	GUM	0.04	**0.94**	0.02	0.04	**0.96**	0.00	0.04	**0.96**	0.00
	GPD	0.78	0.00	**0.22**	0.20	0.00	**0.80**	0.58	0.00	**0.42**

Table 5.1: Proportion of selecting a parametric model using histogram $\tilde{T}$, transformed kernel $\hat{T}$ and standard kernel $\hat{T}^*$ based tail indices for the univariate Fré, Gumbel and GPD target densities, and for sample size $n = 100$. The tail indices are calculated using the L_2 (top rows) and χ^2 (bottom rows) errors. The proportion of correctly selecting the target model is given in bold.

variations are only due to the dependence structure between the random variables which is then modelled in the second step. Beranger and Padoan (2015) review the main families of parametric extremal dependence models and provide the expression of their multivariate distributions and the densities are easily derived from there. We consider some models that are within the class of logistic distributions: the asymmetric negative logistic (Joe, 1990) and the bilogistic (Smith et al., 1990) distributions. The Hüsler-Reiss model (Hüsler and Reiss, 1989) is also considered.

We begin by taking an $m \approx 4000$ sample from each of these three target densities: bilogistic (BIL) with $\alpha = 0.80$ and $\beta = 0.52$, asymmetric negative logistic (ANL) with dependence 1.3 and asymmetry $(0.2, 0.7)$ and Hüsler-Reiss (HR) with dependence 2.4. The bivariate extreme sample is made of an $n = 200$ random selection of the observations that are exceed jointly the threshold $\boldsymbol{u}$ which is set to the marginal sample 90% upper quantiles. For the parametric models, the marginal parameters and the dependence structure are simultaneously estimated, and based on the assumption that the margins are GEV distributed, their maximum likelihood estimates are taken as starting values in the likelihood optimization. The non-parametric estimators are a 2D histogram with a normal scale binwidth (HIS) and a transformation kernel estimator with a plug-in bandwidth (KPI), as we no longer consider the untransformed estimator KPI* due to its poor performance for the univariate simulations. The results from the other bandwidth selectors are not displayed for clarity.

In Figure 5.3, the target density is the solid black contours, the histogram the dashed red, and the kernel estimator the dot-dashed blue. The contour lines indicate the $25, 50, 75$ and 99% highest density level sets in the tail. It is known that 2D qq-plots are not well-defined and consequently are not provided as in Figure 5.1. Unlike qq-plots, our density estimates visualisations don't suffer from this lack of well-definedness in higher dimensions. Furthermore, the good visualisation provided by the kernel estimator provides easy interpretation of the tail behaviour, in contrast to the blocky, discrete nature of the histogram which is unable to reveal clearly this behaviour. The kernel estimator most accurately estimate the tail density of the HR density, with the performance decreasing for the BIL and ANL densities.

In Figure 5.4 the box-plots compare the accuracy of the histogram and kernel estimators to

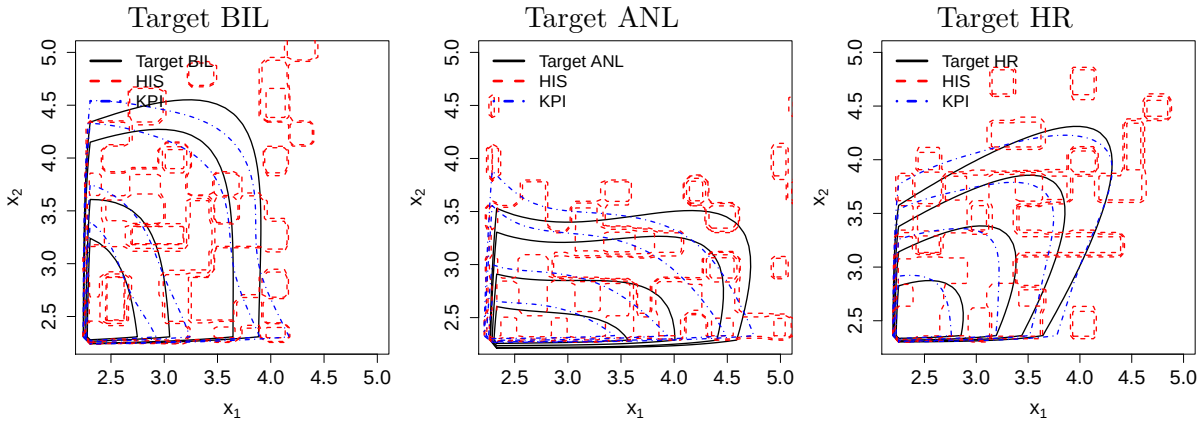

Figure 5.3: Non-parametric estimators of the bivariate tail density when the target density is bilogistic (BIL), asymmetric negative logisitic (ANL) and Hüsler-Reiss (HR). Target density – solid black, histogram with normal scale binwidth (HIS) – dashed red, and kernel estimators with the plug-in (KPI) – dot-dashed blue. Sample size is $n = 200$.

the target parametric estimator for bivariate tail samples drawn from the bilogistic, asymmetric negative logistic and Hüsler-Reiss distributions for both measures of accuracy. As expected the correctly specified parametric estimator of each distribution provides the smallest error, although for the asymmetric negative logistic distribution the plug-in estimator performs equally well and the histogram produces more accurate fits than for other distributions. The kernel estimators provides uniformly lower errors than the histograms over all target densities.

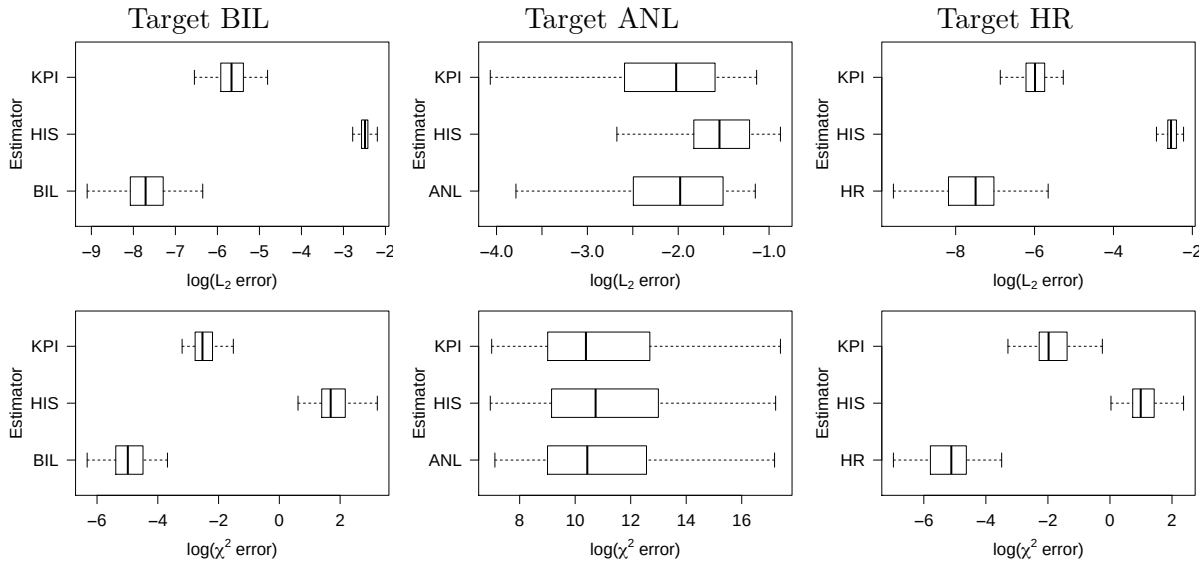

Figure 5.4: Box-plots of the $\log L_2$ (top) and $\log \chi^2$ (bottom) errors for the bivariate bilogistic (BIL), asymmetric negative logistic (ANL), Hüsler-Reiss (HR), histogram (HIS) and bivariate kernel (KPI) tail density estimators. For each panel the bottom box-plot is the parametric estimator associated to the target density. From left to right the target densities are BIL, ANL, HR. Sample size is $n = 200$.

We compute histogram and kernel based tail indices $\tilde{T}$ and $\hat{T}$ with respect to each parametric model, based on the L_2 and χ^2 measures of error. The proportions of parametric models that

these tail indices select are given in Table 5.2, with the correctly selected ones in bold. The bold figures suggest that $\tilde{T}$ slightly outperforms $\hat{T}$ when the bilogistic is the target density. Overall, the L_2 error is more accurate than the χ^2 error in selecting the target model, with the notable exception is the correctly selected ANL models with $\hat{T}$. This is in contrast to the univariate model selection where the two error measures yield similar accuracy. Examining the L_2 model selections more thoroughly, the propensity of $\tilde{T}$ to select the right model more often than $\hat{T}$ when the target density is the asymmetric negative logistic is highlighted. $\hat{T}$ selects more often the bilogistic than the true asymmetric logistic density. However note that when the target is Hüsler-Reiss, $\hat{T}$ selects the right model a large majority of the time (96%).

		$\tilde{T}$ selected model			$\hat{T}$ selected model		
Error	Target	BIL	ANL	HR	BIL	ANL	HR
L_2	BIL	**0.70**	0.23	0.07	**0.61**	0.23	0.16
	ANL	0.15	**0.84**	0.01	0.74	**0.26**	0.00
	HR	0.17	0.07	**0.76**	0.00	0.04	**0.96**
χ^2	BIL	**0.00**	0.12	0.88	**0.51**	0.22	0.27
	ANL	0.38	**0.01**	0.61	0.23	**0.76**	0.01
	HR	0.00	0.01	**0.99**	0.00	0.00	**1.00**

Table 5.2: Proportion of selecting a parametric model using histogram $\tilde{T}$ and kernel $\hat{T}$ based tail indices using the L_2 and χ^2 measures of error for the bilogistic (BIL), asymmetric negative logistic (ANL), Hüsler-Reiss (HR) densities, and for sample size $n = 200$. The proportion of correctly selecting the target model is given in bold.

5.3.3 Analysis of climate models

Perkins et al. (2007, 2013) used histogram density estimators for visualisation and model selection in the extreme value modelling to calculate return levels of minimum and maximum temperatures in Australia, based on data generated from an ensemble of global climate models (GCMs). The models which they considered are the climate models assessed by the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) to investigate changes in temperature extremes. A well-known challenge for these models is to be able to accurately project extreme temperatures Perkins et al. (2007, 2013); Sillmann et al. (2013a). Perkins et al. (2013) consider three skills to assess the quality of the fit of each AR4 model. The first skill principle relies on comparing the means of each model to the observed data, which is an inadequate indicator of the model's ability to simulate extremes Perkins et al. (2007). A second approach was defined by Perkins et al. (2007) and provides a score reflecting the overlap between the observed and modelled p.d.f's. Although it is an improvement it was not designed specifically to compare extreme values. The extension introduced by Perkins et al. (2013) focuses on the tail of the densities where the extreme events are located. This score reflects the discrepancy between two tails where a model perfectly fits the observed data has zero score and increasing scores imply an increasing lack-of-fit of the model to the observed data. Unlike for the simulated parametric models in Section 5.3.1 and 5.3.2 , there is no closed form for the density function $g_i, i = 1, \ldots, M$, to characterise the simulated values generated by the climate

models. Perkins et al. (2013) replace the target density g_i with a histogram $\tilde{g}_i$, based on model generated data, when computing the goodness-of-fit to the histogram of the observed data in Equation (5.4), i.e. $\tilde{T}(\tilde{g}_i)$, to decide the most appropriate model

We extend this analysis by considering a wider and more modern ensemble of global climate models than those in Perkins et al. (2013), as well as exploring alternatives to $\tilde{T}(\tilde{g}_i)$ as the model selection criterion. Here we use $M = 22$ climate models participating in the World Climate Research Programme's 5th phase Coupled Model Intercomparison Project (CMIP5; see Flato et al. (2013)). This ensemble of numerical models currently underpin global and regional climate projections of extremes (e.g. Sillmann et al. (2013b)), and the subset used in this study was based on the availability of daily maximum and minimum temperature data for the historical experiment ($\sim 1860 - 2005$; see Taylor et al. (2012)). Other targeted temperature extreme evaluation studies on the CMIP5 ensemble have found generally well-simulated changes in observed trends of specific indices (e.g. Sillmann et al. (2013a); Flato et al. (2013)), however unlike this study, have not considered the simulation of the underlying distributions.

The grid box that included, or was nearest to Sydney was extracted from each model for both climate variables, and truncated in time to match the observations (see below).

We propose another criterion: $\hat{T}(\hat{g}_i)$ (kernel tail index with respect to a kernel estimator) of the model data. The observed data sample are the daily observed maximum temperatures observed in the Sydney region in Australia from 01/01/1911 to 31/12/2005 yielding to a sample of $m = 34699$ observations. Observations were obtained from the Australian Water Availability Project dataset (Jones et al. (2009)), a gridded product covering all of Australia. The grid box in which Sydney is located was extracted. All AR4 climate models have been run to generate data in this same time frame.

The threshold for the maximum temperatures is the 95% upper quantile $u = 30.98°$C and tail sample size $n = 1727$. Table 5.3 contains the two variants of these tail indices for the maximum temperatures, using both the L_2 and χ^2 errors, with the three minimal values given in bold. From the table, the PMI-ESM-LR and MPI-ESM-MR models are selected by both $\tilde{T}(\tilde{g}_i)$ and $\hat{T}(\hat{g}_i)$ for each the measure of error. The third model selected by the histogram indices is CMCC-CESM whereas the CanESM2 and CNRM-CMS models are simultaneously selected by the kernel indices. It is also important to note that for all selection criteria, the ACESS1-3, NorESM1-M, IPSL-CM5* and MIROC-ESM* models produce inadequate fits.

Figure 5.5 plots the tail density estimators for the selected CMCC-CESM, HadGEM2-ES, PMI-ESM-LR, MPI-ES-MR model data and observed data. The histograms are the solid lines, kernel estimators the dashed lines. The observed data are black, and the simulated model data blue. For PMI-ESM-LR and MPI-ES-MR models, both the histogram and kernel estimators of the simulated data closely follow both the histograms and kernel estimator of the observed data, so in these cases, the choice of the estimator of the observed data plays a small role. The qq-plots confirm that the PMI-ESM-LR and MPI-ES-MR climate models most closely fit the observed data, whereas the HadGEM2-ES and CMCC-CESM models fit the observed data less well for maximum daily temperatures less than 38°C.

For bivariate data, we consider the pairs of maximum and minimum temperatures over

	Model selection criterion			
	$\tilde{T}(\tilde{g})$		$\hat{T}(\hat{g})$	
	Error measure		Error measure	
Model	L_2	χ^2	L_2	χ^2
ACCESS1-0	0.073	0.106	0.076	0.075
ACCESS1-3	0.299	2.047	0.296	6.344
CanESM2	0.058	0.081	0.040	**0.015**
CMCC-CESM	**0.046**	**0.065**	0.033	0.026
CMCC-CM	0.052	0.067	0.034	0.019
CMCC-CMS	0.063	0.096	0.059	0.046
CNRM-CMS	0.069	0.081	0.033	**0.015**
CSIRO-Mk3-6-0	0.061	0.083	0.055	0.040
HadCM3	0.064	0.095	0.065	0.069
HadGEM2-CC	0.055	0.086	0.034	0.028
HadGEM2-ES	0.058	0.076	**0.030**	0.017
IPSL-CM5A-LR	0.164	0.274	0.146	0.164
IPSL-CM5A-MR	0.108	0.130	0.085	0.056
IPSL-CM5B-LR	0.153	0.243	0.139	0.163
MIROC-ESM	0.189	0.483	0.176	0.303
MIROC-ESM-ECHM	0.184	0.458	0.167	0.283
MIROC5	0.066	0.078	0.046	0.019
PMI-ESM-LR	**0.045**	**0.058**	**0.028**	**0.005**
MPI-ESM-MR	**0.045**	**0.057**	**0.024**	**0.005**
MPI-ESM-P	0.077	0.104	0.072	0.044
MRI-CGCM3	0.075	0.121	0.078	0.113
NorESM1-M	0.124	0.576	0.072	0.138

Table 5.3: Histogram $\tilde{T}$ and kernel $\hat{T}$ based indices respectively for histogram $\tilde{g}$ and kernel $\hat{g}$ estimators of the observed data, using the L_2 and χ^2 errors, for the 24 AR4 climate models for maximum temperatures. The selected models with the three minimal criteria values are in bold.

the same time period, in order to investigate which of the climate models can predict a joint increase. The threshold for the maximum temperatures are the 90% magrinal upper quantiles $\boldsymbol{u} = (28.77°\text{C}, 18.07°\text{C})$ and tail sample size is size $n = 1387$. Table 5.4 contains the two variants of the tail indices based on both the L_2 and χ^2 errors for the (maximum, minimum) temperatures, with the three minimal values given in bold. The CNRM-CMS model is selected by all four indices and systematically provides the lowest values. HadCM3 has also been identified by $\tilde{T}(\hat{g})$ and $\hat{T}(\hat{g})$ as a good model for the prediction of extreme minimum and maximum temperatures. Note that the NorESM1-M model which performed poorly for the marginal univariate maximum temperature analysis above, becomes a viable model when considering the joint (maximum, minimum) extremes.

Figure 5.6 illustrates the plots of the bivariate tail density estimators for the selected CNRM-CMS and HadCM3 models. Similarly to Figure 5.5, histograms are the solid lines, kernel estimators the dashed lines. The observed data are black, and the simulated model data blue. The contour lines indicate the $25, 50, 75$ and 99% highest density level sets in the bivariate tail of the kernel estimator of the observed data. Both GCMs models appear to simulate extremes

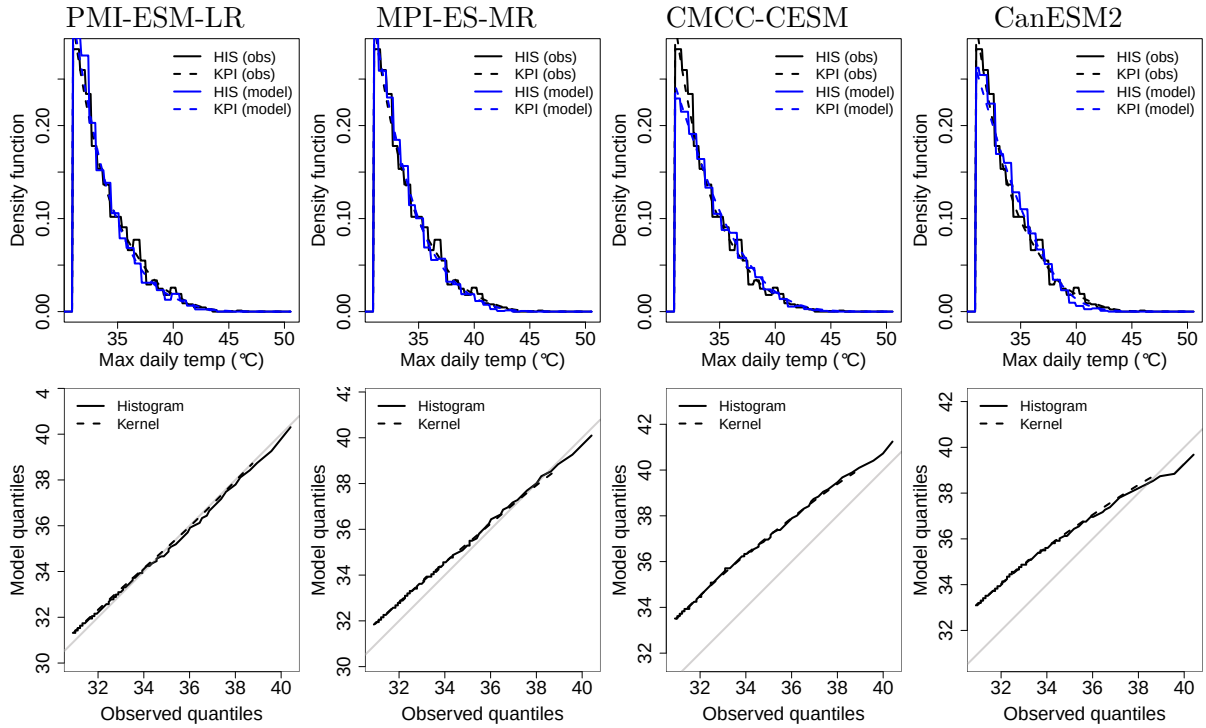

Figure 5.5: Histogram and kernel estimators of the tail densities for the selected AR4 models: CMCC-CM, HadGEM2-ES, PMI-ESM-LR, MPI-ES-MR. Histograms estimators (HIS) are the solid lines and kernel plug-in estimators (KPI) the dashed lines. Observed data (obs) is black and modelled data (model) is blue.

of the maximum temperatures rather well, however the magnitude of minimum temperatures are under-estimated (i.e. they are too warm). This indicates two possibilities: 1) that the kernel estimators need to be further refined at the boundary; 2) physical parameters in the GCMs need to be revised for a more realistic simulation of minimum temperature. While minimum temperatures are somewhat easier for a climate model to simulate over maximum temperature (Perkins et al. (2007)), the under-representation of the observed temperature distribution is a well-known issue in GCMs, which is at least in part explained by their coarse resolution (Seneviratne et al. (2012)). For example, dynamically downscaled regional climate models that are run at finer resolutions for a limited spatial domain can offer some improvement in the simulation of extreme temperatures (Seneviratne et al. (2012); Vautard et al. (2013); Perkins et al. (2014)).

5.4 Discussion

We have introduced non-parametric kernel estimators for the analysis of the tail density for univariate and multivariate extreme values data. Our contribution consists of applying a transformation kernel estimation which accounts for the boundedness and skewness of the tail sample to the tail sample only. In addition to excellent visualisations of this tail density estimator, we included it in a model selection procedure based on tail index which measures its

	Model selection criterion			
	$\tilde{T}(\tilde{g})$		$\hat{T}(\hat{g})$	
	Error measure		Error measure	
Model	L_2	χ^2	L_2	χ^2
ACCESS1-0	0.0486	8.4892	0.0018	7733.3500e0
ACCESS1-3	0.0282	6.0338	0.0036	1.8028e7
CanESM2	0.0220	4.2660	0.0018	6.6574e0
CMCC-CESM	0.0751	11.3020	0.0018	1.8374e4
CMCC-CM	0.0067	1.2834	**0.0010**	8313.6940e0
CMCC-CMS	0.1477	23.3683	0.0028	369.2623e0
CNRM-CMS	**0.0034**	**0.6459**	**0.0005**	**0.3955e0**
CSIRO-Mk3-6-0	0.0484	8.4685	0.0018	1528.0070e0
HadCM3	**0.0042**	**0.9516**	**0.0007**	9.4845e4
HadGEM2-CC	0.0354	5.4054	0.0016	2.0869e0
HadGEM2-ES	0.0333	5.5201	0.0013	4.5137e0
IPSL-CM5A-LR	0.0114	2.2945	0.0015	2.7308e0
IPSL-CM5A-MR	0.0176	3.3012	0.0014	4.7341e0
IPSL-CM5B-LR	0.0163	2.5026	0.0014	**0.4354e0**
MIROC-ESM	0.0259	5.3681	0.0014	9.3011e0
MIROC-ESM-ECHM	0.0335	7.0436	0.0083	27.6267e0
MIROC5	0.0481	7.4235	0.0044	241.9099e0
PMI-ESM-LR	0.0155	2.4104	0.0011	3293.9260e0
MPI-ESM-MR	0.0315	4.9851	0.0015	4858.4120e0
MPI-ESM-P	0.0178	2.4501	0.0011	4.4354e0
MRI-CGCM3	0.0554	11.1943	0.0024	1.3437e5
NorESM1-M	**0.0063**	**0.7126**	0.0012	**0.1151e0**

Table 5.4: Histogram $\tilde{T}$ and kernel $\hat{T}$ based indices for histogram $\tilde{g}$ and kernel $\hat{g}$ estimators of the observed data, for the 24 AR4 climate models for (maximum, minimum) temperatures. The selected models with the three minimal criteria values are in bold.

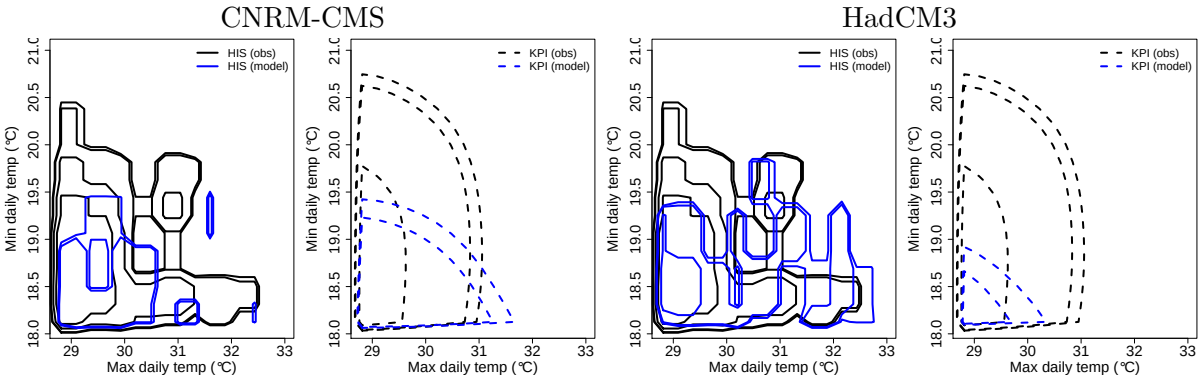

Figure 5.6: Histogram and kernel estimators of the tail densities for the selected AR4 models: CNRM.CMS, HadCM3. Histograms estimators (HIS) are the solid lines and kernel plug-in estimators (KPI) the dashed lines. Observed data (obs) is black and modelled data (model) is blue.

discrepancy to the kernel estimators of a suite of parametric model tail densities. We verified the performance of these analyses for simulated and climate experimental data.

Future perspectives include refining the kernel estimator at the boundary to further improve the visualisation as well as removing our heuristic double truncation to compute the tail index, and to evaluate the limiting distribution of the tail index which would allow us to make probabilistic statements about the observed tail index values. For the climate data, as the climate models do not allow for probabilistic predictions, their analysis usually proceeds to another round of model fitting, this time from a suite of statistical models. Once this is selected, we can take advantage of its statistical properties to provide probabilistic interpretations of predictions of extreme climactic events.

Summary and Discussion

In this section, we first state the main findings of this thesis, discuss their advantages but also their potential drawbacks or limitations. We then put the emphasis on possible extensions and future research ideas.

Chapter 1 provides an original contribution in the following sense. High-dimensional expressions of the angular densities and exponent functions of the main extremal dependence models in the multivariate setup are given. These are helpful in order to carry out parameter estimation and to compute summaries of the dependence structure. We use the expression of the exponent function of the extremal-t model given by Nikoloulopoulos et al. (2009) to derive its angular density on all the sub-spaces of the simplex. The angular density derived from this model has the particularity to put mass on all the subsets of the simplex. The expression of the density on the interior and corners of the simplex are given while expression (4.19) enables their derivation on the other sub-spaces. We performed four-dimensional analyses of pollution data for many different extremal dependence models using both the likelihood based approach and the Bayesian approach described by Cooley et al. (2010) and Sabourin et al. (2013). Probabilities of jointly exceeding some fixed high threshold are computed and used to estimate high-dimensional return levels. They provide the estimated (return) period for combinations of air pollutants to reach some high levels. The pros and cons of the approximate likelihood method are further investigated and discussed in Chapter 4.

The main argument given in Chapter 3 is that any high-dimensional marginal distribution of the Brown-Resnick can be analytically derived. They are given in the special case introduced by de Haan and Pereira (2006), further applied to real data by Buishand et al. (2008), where a general spatial dependence parameter is introduced and a bivariate construction is considered. The result allows to fit the full likelihood estimation procedure to the model and not having to use approximation methods. Another possibility is to keep approximating the likelihood function by high degrees of the composite likelihood. However both of these suggestions lead to an increase of the computational load and a trade-off has to be made. Unfortunately these results were derived in parallel by Huser and Davison (2013) for the generalized Brown-Resnick model and for which the efficiency of different composite likelihood estimators were compared. This highlights the fact that spatial extremes is a very active research field motivated by the growing interest in the projection of extreme events in areas like meteorology, for instance.

Chapter 4 provides a simple definition of a non-stationary non-normal process based on the additive stochastic representation of skew-normal variables. With a covariance function de-

pending on the spatial lag these processes are not strictly stationary, however, weak stationarity occurs if the slant function is a fixed function of space. Moreover the joint tail dependence model from Ledford and Tawn (1996) is used to show asymptotic independence for the skew-normal distribution. We introduce the extremal skew-t process, a new max-stable process based on an underlying skew-normal process and constructed using the spectral representation. The main advantage of the extremal skew-t model is that, unlike most max-stable models, it exhibits non-stationarity. This is due to the presence of the slant function in the dependence structure. It allows to derive also non-stationary extremal coefficient functions which vary depending on the spatial location. The finite-dimensional distributions are given as a function of non-central extended skew-t cdfs. We define such distribution, not available before, which includes the standard extended skew-t distribution defined by Arellano-Valle and Genton (2010). Additionally, we show that the same dependence structure can be obtained when considering the limiting distribution of appropriately rescaled Skew-t random variables.

Furthermore, setting the slant parameters to zero reduces to the stationary Extremal-t model. Similarly to Chapter 1, the expression of the angular density of both the interior and corners of the simplex are given for arbitrary dimensions. The mass on other sub-spaces of the simplex is also derived for the three and four-dimensional angular densities and thus the approximate likelihood approach is considered for parameter estimation. It highlights the fact that the density on all the subsets of the simplex are required for an accurate estimation. This is in agreement with Sabourin et al. (2013) who stated that likelihood based on the angular density needs to be computed by differentiation of order d (the dimension). Although approximate likelihood might be less computationally demanding than composite likelihood methods, it has the drawback of adding extra arbitrariness due to the allocation of a data point to a specific subset of the simplex. In the spatial context, considering a simple correlation structure and using the composite likelihood estimation method, we show that there seems to be a gain in efficiency, when using triplewise rather than pairwise composite likelihood, for smoother processes.

Finally Chapter 5 presents a non-parametric estimation method for tail-densities using kernel estimators. This kernel-based method provides a good graphical tool for exploratory data analysis. It focuses on the tail sample only which is transformed to be unbounded and to enjoy the good properties of kernel density estimators. We show that our estimator has the same good asymptotic properties as usual kernel density estimators in both univariate and multivariate contexts. Moreover we demonstrate that our method outperforms other non-parametric estimators such as standard kernel density estimators and histograms. Our estimator is then used as a surrogate for the true underlying density and tail indices are computed for different parametric models. An illustration of this method's good performances for model selection is given for datasets generated from different univariate and bivariate extreme value distributions. Additionally it is shown to be more consistent than the histogram based tail index introduced in Perkins et al. (2013). Finally a real data example showcases a scenario where our method is required. The ability of several general circulation models (GCMs) to simulate extreme minimal and maximal temperatures in the region of Sydney, Australia, is investigated. Again the kernel based tail index exhibits more consistency than the histogram based tail index and the good

visualisation it provides is highlighted.

As mentioned above, the simulation studies conducted in Chapter 1 and developed deeper in Chapter 4 highlight some weakness of the approximate likelihood method. This promising method requires further attention. Selecting a rule to decide whether a point belong to the interior of the simplex or to another subspace of the simplex, can be established easily up to three dimensions but is much harder in higher dimensions due to its representation. The rules taken into account in Chapter 1 and 4 were homogeneous in the way they were identical for each subspace. This lead to accurate results for three-dimensional data but keeping the same rule as the dimension increases seems unlikely. Hence there is a strong motivation to define theoretical results about the selection of these rule would strengthened the use of this method which is still relatively untouched in the literature. An advantage of this method is that it reduces the computational burden that is faced by full likelihood estimation or composite likelihood (triplewise or higher) when there is a large number of spatial locations. This leads to a second potential improvement of the approximate likelihood method. The possibility to derive of a "composite approximate" likelihood method which could lead to much faster estimation and bypass the decision rule dilemma (if restricted to triplewise composite approximate), will be investigated.

Chapter 4 opens other questions. First, only a power-exponential correlation function has been considered in the spatial setting. Section 2.6 introduced various types of continuous functions. Their effect on the dependence structure of the extremal skew-t should be studied. Moreover, adding some extra conditions can lead to specific types of correlation structures that allow for ergodic or mixing processes. Furthermore, it is important to study different correlation structures in order to highlight the ones that are easy to fit and simple to use in practice.

Additionally, the model selection method given in Chapter 5 and based on the kernel-based tail index, didn't take the problem of bias at the boundary of the support into account. Because the focus is on the observations in the far tail, a first solution to bypass this problem is to apply a second truncation before calculating the tail indices in order to reduce the bias. Also, transformation functions to an unbounded support are not unique and even though the one considered in Chapter 5 is the most natural, other transformations can reduce bumps at the boundary. Furthermore, Geenens (2014) proposed an efficient method to reduce bias by combining a probit transformation and a local likelihood density estimation. These solutions to a more accurate estimate and to a better model selection, will be investigated in further research.

Finally, to conclude, the current context (global warming, financial crises, etc) make the population more aware and worried about the possibility of extreme events. As the challenges are expanding so are researchers interests in this field of statistics.

Appendix A

Convergence in law of a simple point process to a Poisson point process

A.1 Basics

Skorohod's theorem (Billingsley, 1971) states that for a simple point process P_n to converge weakly to a Poisson point process P, there must exist a sequence of random elements defined on the uniform probability space, with intensity measure the Lebesgue measure, such that this sequence is equal in distribution to the simple point process P_n and converges almost surely. Weak convergence means that actual realizations of the sequence of random variables do not converge at all and oscillate widely. Skorohod's theorem, however, proves the existence of an equally likely sequence of random variables that converge with probability one.

Let P_n denote a simple point process on $\mathbb{R}_+ \times \mathbb{R}_+$ with points $\left\{\left(\frac{k}{n}, nX_k\right), k \geq 1\right\}$, where $X_1, \ldots, X_n$ are i.i.d. random variable uniformly distributed over $[0, 1]$. Our aim is to prove that P_n converges weakly to a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measure the Lebesgue measure.

In order to do so we will first define a sequence $\tilde{P}_n$ that is equal in distribution to P_n and then we will prove that it converges almost surely. Denote by Q_n the restriction of P to $\mathbb{R}_+ \times [0, n]$ whose points are $\{(\tau_k(n), y_k(n)), k \geq 1\}$ where the $\tau_k(n), k \geq 1$ are ordered. Abscissas $\tau_k(n)$ correspond to the projection of the points of the Poisson point process on the x-axis, forming a homogeneous Poisson point process on the $[0, \infty)$. Furthermore the Lebesgue measure is a product measure, meaning that the $y_k(n)$ are random variables uniformly distributed on $[0, n]$. It is then straightforward to see that if $\tilde{P}_n$ defines a sequence whose points are $\left\{\left(\frac{k}{n}, y_k(n)\right), k \geq 1\right\}$, then $\tilde{P}_n \stackrel{d}{=} P_n$ and we are left with proving its almost convergence.

First, remark that $\tilde{P}_n$ can be obtained from Q_n through through the horizontal shift, or random mapping, $\frac{k}{n} - \tau_k(n)$. Additionally, by construction Q_n converges to P almost surely as n get large and thus it suffices to show that the shift in the first coordinate of Q_n converges to zero as n increases to obtained almost sure convergence of $\tilde{P}_n$. In other words we only need to prove the convergence of the random mapping producing $\tilde{P}_n$ from Q_n to the identity mapping as

n increases. Without loss of generality we will restrict ourself to the analysis of the convergence on $(0, S) \times \mathbb{R}$. For $S > 0$, we aim to show

$$\sup_{0 \leq k/n \leq S} \left| \frac{k}{n} - \tau_k(n) \right| \xrightarrow{a.s.} 0, \text{ as } n \to \infty,$$

and subsituting k/n by x, this is equivalent to

$$\sup_{0 \leq x \leq S} |x - \tau_{[nx]}(n)| \xrightarrow{a.s.} 0, \text{ as } n \to \infty.$$

The sequence $\{\tau_k(n), k \geq 1\}$ is strictly increasing in k and so the function $\tau_{[nx]}(n)$ for $0 \leq x \leq S$ is monotone for all n. Hence we can consider the limit of the inverse of $f(x) = \tau_{[nx]}(n)$ is given by $f^{-1}(x) = n^{-1} \max_k \{\tau_k(n) \leq x\}$. The maximum term can be seen as the number of points of the Poisson point process P that fall in the region $(0, x] \times (0, n]$. Denote by $N_k(x)$ the number of points of P that fall in $(0, x] \times (k-1, k]$ for $k = 1, \ldots, n$, then this yields tp

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} N_i(x) = 0,$$

by the law of large numbers. This proves that the horizontal shift tends to zero as n get large and consequently $\tilde{P}_n$ converges almost surely to P.

A.2 In the context of Extremes

Let P_n now denote a simple point process on $\mathbb{R}_+ \times \mathbb{R}_+^d$ with points

$$\left\{ \left(\frac{k}{n}, \frac{X_{k1} - b_{n1}}{a_{n1}}, \ldots, \frac{X_{kd} - b_{nd}}{a_{nd}} \right), k \geq 1 \right\},$$

where $X_1, \ldots, X_n$ are i.i.d. random variable in $\mathbb{R}^d$ with distribution F and margins $F_i, i = 1, \ldots, d$. The positive constant a_{ni} and $b_{ni}, i = 1, \ldots, d$, are such that F is in the maximum domain of attraction of some multivariate extreme value distribution G. It is well known that if U_i denotes the inverse function of $1/(1 - F_i)$ i.e $U_i(n) = F_i^{\leftarrow}(1 - 1/n)$, then

$$\lim_{n \to \infty} \frac{U_i(nx_i) - b_i(n)}{a_i(n)} = \left(\frac{1}{-\log G_i} \right)^{\leftarrow} (x_i), \; i = 1, \ldots, n, \tag{A.1}$$

where G_i is a univariate extreme value distribution. The same restriction on the points of P leads to the sequence Q_n and from there a straightforward choice of the sequence $\tilde{P}_n$ is to let its points be

$$\left\{ \left(\frac{k}{n}, \frac{U_1(n/y_{k1}(n) - b_{n1})}{a_{n1}}, \ldots, \frac{U_d(n/y_{kd}(n) - b_{nd})}{a_{nd}} \right), k \geq 1 \right\},$$

where $a_i(n) = a_{ni}$ and $b_i(n) = b_{ni}$ and the sequence of i.i.d. random variables $y_{ki}(n)$ are uniformly distributed on $[0, n]$. Consequently using the definition of U_i, we can conclude that the sequence $\{(U_1(n/y_{k1}(n)), \ldots, U_d(n/y_{kd}(n))), k \geq 1\}$ has distribution function F and thus

$\tilde{P}_n \stackrel{d}{=} P_n$. The random shift to go from Q_n to $\tilde{P}_n$ is similar to the one given in the previous section but the transformation of the second parameter is no longer random. However, using the same steps as previously, we are able to show that $\tilde{P}_n$ converges almost surely to a Poisson point process P' obtained by applying the transformation given by the RHS of (A.1) to the second up to the last coordinate of P. The transformation theory for Poisson point processes assures that the intensity measure of P' is $\mathrm{d}t \times \mathrm{d}\nu$ evaluated at $[0,t] \times (\boldsymbol{x}, \boldsymbol{\infty}), \boldsymbol{x} \in \mathbb{R}^d$ which is $-t \log G(\boldsymbol{x})$. Applying the mapping theorem from Billingsley (1968, Theorem 5.1), the convergence of the point process with points $\left\{ \left(\frac{X_{k1}-b_{n1}}{a_{n1}}, \ldots, \frac{X_{kd}-b_{nd}}{a_{nd}} \right), k \geq 1 \right\}$ to a Poisson point process on $\mathbb{R}$ with intensity $\mathrm{d}\nu$ where $\nu(\boldsymbol{x}, \boldsymbol{\infty}) = -\log G(\boldsymbol{x})$.

Appendix B

Construction of the Brown-Resnick model

In this appendix, the model called Brown-Resnick in Section 2.6.3 is proved to be obtained by considering the convergence in distribution of rescaled i.i.d. Ornstein-Ulhenbeck processes and the conditions of Theorem 2.5 are verified. In a first section we need to define the norming constants a_n and b_n when studying the limiting distribution of the rescaled maxima of Gaussian random variables. This result is then taken into account in a second section where the convergence of the appropriately rescaled Brownian motions. Finally the extension to Ornstein-Ulhenbeck processes is established, following closely the lines of the Brownian case.

B.1 Normalizing constants for a Normal distribution to be in the maximum domain of attraction of an extreme value distribution

Let $X_1, X_2, \ldots$ be i.i.d. normally distributed random variables with mean 0 and variance σ^2 and denote their distribution by F and their maxima $M_n = \max(X_1, \ldots, X_n)$. Assume that, for all $n \in \mathbb{N}$, there exists constants $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x),$$

where $G(x)$ is a type I extreme value distribution.

Using Mill's ratio we have $\bar{\Phi}(x) \sim \phi(x)/x$ and assume the right hand side to be the tail of some H, i.e. Φ and H are tail equivalent. According to Embrechts et al. (1997, Proposition 3.3.28), if Φ and H have the same right endpoint $x_\Phi = x_H$ and Φ is in the maximum domain of attraction of G (as defined above) then both distributions have the same norming constants if and only if they are tail equivalent. Moreover it is easy to show that Φ is a Von Mises function with auxiliary function $a(x) = \bar{\Phi}(x)/\phi(x) \sim x^{-1}$ and hence the norming constants can be given

by

$$a_n = a(b_n) \quad \text{and} \quad b_n = H^{\leftarrow}(1 - n^{-1}),$$

which is equivalent to

$$a_n = b_n^{-1} \quad \text{and} \quad \bar{H}(b_n) = n^{-1}.$$

Using the tail equivalence, the second equation reduces to $\ln b_n + \frac{1}{2}\ln 2\pi + \frac{1}{2}b_n^2 = \ln n$ and using a Taylor expansion it yields

$$b_n = (2\ln n)^{1/2} - \frac{\ln\ln n + \ln 4\pi}{2(2\ln n)^{1/2}} + o\left((\ln n)^{-1/2}\right).$$

An appropriate choice for a_n is then $(2\ln n)^{-1/2}$.

B.2 Convergence in distribution of the maximum of Brownian motions

Let us now consider $\{W_i(t)\}_{i\geq 1, t\geq 0}$ a sequence of i.i.d. standard Brownian motions with initial condition $W_i(0) = 0$ and zero drift. We will show the convergence in distribution of the standardized maxima of the W_i's in the neighborhood of some value $t_0 > 0$. The normalization introduced above allows us to define the following process

$$\xi_{n,i}(s) = b_n t_0^{-1/2}\left(1 + sb_n^{-2}\right)^{-1/2}\left\{W_i\left(t_0\left(1 + sb_n^{-2}\right)\right) - b_n t_0^{1/2}\left(1 + sb_n^{-2}\right)^{1/2}\right\}, i \geq 1$$

for $s \geq 0$. We know that $M_n(s) := \max_{1\leq i\leq n}\xi_{n,i}(s)$ converges in distribution to Y with type I extreme value distribution. Furthermore $\{\sqrt{t_0}W(s/t_0)\}, s \geq 0$ is standard Brownian motion for any fixed t_0 we choose $t_0 = 1$. Without loss of generality, we will restrict ourself to prove the convergence of the point process to a Poisson point process on C[0, 1].

Use the decomposition $W_i(1 + sb_n^{-2}) = W_i(1) + b_n^{-1}W_i^*(s)$, where $\{W_i^*, i \geq 1\}$ are i.i.d standard Brownian motions. Note also that $\{W_i^*, i \geq 1\}$ and $\{(W_i(s), 0 \leq s \leq 1), i \geq 1\}$ are independent. Then noting that $b_n \uparrow \infty$ and when $n \to \infty$

$$\left(1 + sb_n^{-2}\right)^{1/2} = 1 + \frac{1}{2}sb_n^{-2} + o\left(b_n^{-2}\right).$$

Then we can write that, uniformly over $s \in [0, 1]$, we have,

$$\begin{aligned}\xi_{n,i}(s) &= b_n\left\{W_i(1) + b_n^{-1}W_i^*(s) - b_n - \frac{1}{2}sb_n^{-1} + o\left(b_n^{-1}\right)\right\}\left(1 + O\left(b_n^{-2}\right)\right)\\ &= b_n(W_i(1) - b_n) + W_i^*(s) - \frac{s}{2} + o\left(b_n^{-1}\right).\end{aligned}$$

Define

$$\xi^*(s) = b_n\left(W_i(1) - b_n\right) + W_i^{**}(s)$$

where for $s \in [0, 1]$, $\{W_i^{**}(s), i \geq 1\} = \{W_i^*(s) - s/2, i \geq 1\}$ are standard Brownian motions

with drift $-1/2$, such that we have uniformly over $1 \leq i \leq n$

$$\lim_{n\to\infty} |\xi_{n,i} - \xi^*_{n,i}| = 0.$$

Hence, if suffices to show the convergence of ξ^* in order to establish the convergence in distribution of ξ.

It can be summarized by saying that for any Borelian $\mathcal{A} \in \mathbb{R}$, $\#\{\xi_{n,i} \in \mathcal{A}\}$ is equivalent to $\#\{\xi^*_{n,i} \in \mathcal{A}\}$, where $\#\{\cdot \in \mathcal{A}\} = \sum_{1\geq i\geq n} \mathbb{1}\{\cdot \in \mathcal{A}\}$. We will now prove that

$$\#\{\xi^*_{n,i} \in \mathcal{A}\} \xrightarrow{\mathcal{D}} \#\{T_k + W^{**}_k \in \mathcal{A}\}$$

where T_k are the realizations of a Poisson point process on $(0, \infty]$ with intensity $e^{-z}\mathrm{d}z$.

Using the mapping theorem from Billingsley (1968, Theorem 5.1), we can focus on showing that, for any Borelian $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \in \mathbb{R} \times \mathbb{R}$, $\#\{(b_n(W_i(1) - b_n), W^{**}_i) \in \mathcal{A}\}$ converges to $\#\{(T_k, W^{**}_k) \in \mathcal{A}\}$. If we consider the random variables

$$X = \#\{(b_n(W_i(1) - b_n), W^{**}_i) \in \mathcal{A}\} \quad \text{and} \quad Y = \#\{(T_k, W^{**}_k) \in \mathcal{A}\},$$

it suffices to show that $\Pr\{X = p\}$ converges to $\Pr\{Y = p\}$, for $p = 1, 2, \ldots$. Because of independence between $\{b_n(W_i(1) - b_n)\}$, $\{W^{**}_i\}$ and $\{T_k\}$, we have

$$\begin{aligned}\Pr\{X = p\} &= \Pr\left\{(b_n[W_p(1) - b_n], W^{**}_1) \in \mathcal{A} \cap \cdots \cap \left(b_n[W_p(1) - b_n], W^{**}_p\right) \in \mathcal{A}\right\} \\ &= \prod_{j=1}^{p} \Pr\{b_n(W_j(1) - b_n) \in \mathcal{A}_1\} \prod_{j=1}^{p} \Pr\{W^{**}_j \in \mathcal{A}_2\}\end{aligned}$$

and

$$\begin{aligned}\Pr\{Y = p\} &= \Pr\left\{(T_1, W^{**}_1) \in \mathcal{A} \cap \cdots \cap \left(T_p, W^{**}_p\right) \in \mathcal{A}\right\} \\ &= \prod_{j=1}^{p} \Pr\{T_j \in \mathcal{A}_1\} \prod_{j=1}^{p} \Pr\{W^{**}_j \in \mathcal{A}_2\}.\end{aligned}$$

Thus it suffices to prove the convergence of $\#\{b_n(W_i(1) - b_n) \in \mathcal{A}_1\}$ to $\#\{T_k \in \mathcal{A}_1\}$ for any Borelian $\mathcal{A}_1 \in \mathbb{R}$.

Note that $\{b_n(W_i(1) - b_n)\}$ fulfills the conditions to be convergent in distribution to a double exponential distribution (extreme value distribution of type I with $\mu = 0$ and $\sigma = 1$).

It is straight forward that the conditions are met to apply the results on the convergence of point processes given in Appendix A which lead to

$$\{b_n(W_i(1) - b_n)\} \xrightarrow{\mathcal{D}} \{T_k\}, \text{ as } n \to \infty,$$

where the T_k is a Poisson random measure with mean measure $e^{-t}\mathrm{d}t$ on $(0, \infty]$. Finally, applying

one more time the mapping theorem from Billingsley (1968, Theorem 5.1) to $\xi_{n,i}(s)$ we get

$$\xi_n(s) := \max_{1\leq i} \xi_{n,s}(s) \xrightarrow{\mathcal{D}} \sup_k \{T_k + W_k^{**}\} := \xi.$$

B.3 Convergence in distribution of the maximum of Ornstein-Ulhenbeck processes

A possible extension of the convergence result given in the previous section appears when $\{X(s)\}_{s\in\mathbb{R}}$ is taken to be a zero mean Ornstein-Ulhenbeck process, with variance $\sigma^2 = 1$ and correlation function $\exp(-\alpha|t|)$ where $t \in \mathbb{R}$ and $\alpha = 1/2$, defined by

$$X(s) = \int_{-\infty}^{s} e^{-(s-u)/2} \mathrm{d}W(u)$$

for all $s \in \mathbb{R}$ and W a Brownian motion on $(-\infty, \infty)$.

The proof of the convergence of the maxima of Ornstein-Ulhenbeck processes will follow the lines of the Brownian motion case. According to the justification provided in de Haan and Ferreira (2006), in order to create more dependence we now consider

$$\xi_{n,i}(s) = b_n \left(X_i \left(\frac{s}{b_n^2} \right) - b_n \right), i \geq 1 \tag{B.1}$$

Focusing on the convergence of $\{\max_{1\leq i\leq n} \xi_{n,i}(s)\}$ and re-writing the process as

$$\begin{aligned} X(s) &= \int_{-\infty}^{s} e^{-(s-u)/2} \mathrm{d}W(u) \\ &= e^{-s/2} \int_{-\infty}^{s} e^{u/2} \mathrm{d}W(u) \\ &= e^{-s/2} \left[\int_{-\infty}^{0} e^{u/2} \mathrm{d}W(u) + \int_{0}^{s} e^{u/2} \mathrm{d}W(u) \right] \\ &= e^{-s/2} \left(X(0) + \int_{0}^{s} e^{u/2} \mathrm{d}W(u) \right). \end{aligned} \tag{B.2}$$

Then combining (B.1) and (B.2) we obtain

$$\begin{aligned} \xi_{n,i}(s) &= b_n \left[e^{-s/2b_n^2} \left(X(0) + \int_0^{s/b_n^2} e^{u/2} \mathrm{d}W(u) \right) - b_n \right] \\ &= e^{-s/2b_n^2} \left[b_n \left(X_i(0) + \int_0^{s/b_n^2} e^{u/2} \mathrm{d}W(u) \right) - b_n^2 e^{s/2b_n^2} \right] \\ &= e^{-s/2b_n^2} \left[b_n (X_i(0) - b_n) + b_n \int_0^{s/b_n^2} e^{u/2} \mathrm{d}W(u) + \left(1 - e^{s/2b_n^2} \right) b_n^2 \right]. \end{aligned}$$

Noting that for an arbitrary s_0 such that $-s_0 \leq s \leq s_0$, Taylor series give $e^{-s/b_n^2} = 1 + O(1/b_n^2)$,

then we have

$$\begin{aligned} b_n \int_0^{s/b_n^2} e^{u/2} \mathrm{d}W(u) &= b_n \left[e^{u/2} W(u) \right]_0^{s/b_n^2} \\ &= b_n e^{s/2b_n^2} W\left(s/b_n^2\right) \\ &= b_n \left(1 + O\left(1/b_n^2\right)\right) W\left(s\ b_n^2\right). \end{aligned} \tag{B.3}$$

Taylor series also give $e^{s/b_n^2} = 1 + s/2b_n^2 + 0\left(1/b_n^4\right)$ for $-s_0 \leq s \leq s_0$ which leads to

$$\begin{aligned} \left(1 - e^{s/2b_n^2}\right) b_n^2 &= \left(-s/2b_n^2 + O\left(1/b_n^4\right)\right) b_n^2 \\ &= -s/2 + O\left(1/b_n^2\right), \end{aligned} \tag{B.4}$$

and defining the standard Brownian motion W^* by $W^*(s) := b_n W\left(s/b_n^2\right)$ we can write

$$\xi_{n,i}(s) = b_n(X_i(0) - b_n) + W_i^*(s) - s/2 + O\left(b_n^{-2}\right)$$

Following the steps of Section B.2, it is straightforward to establish the convergence:

$$\xi_n(s) := \max_{1 \leq i \leq n} \xi_{i,n}(s) \xrightarrow{\mathcal{D}} \sup_k \left\{ T_k + W_k^{**}(s) \right\},$$

where T_k are the realizations of a Poisson point process on $(0, \infty]$ with intensity $e^{-z}\mathrm{d}z$ and the margins are assumed to be standard Gumbel.

One may choose to use the transformation $x \mapsto e^x$ to unit Fréchet margins with the limiting process then being

$$\sup_k \left\{ Z_k e^{W_i^*(s) - s/2} \right\},$$

where $\{Z_k\}_{1 \leq k}$ are realizations of a Poisson random measure with intensity $z^{-2}\mathrm{d}z$ on $(0, \infty]$.

To finish we can prove that for $s \in \mathbb{R}$, $Y(s) = e^{W(s) - |s|/2}$ does verify the assumption of integrability $\mathbb{E}[\sup_{s \in \mathbb{S}} Y(s)] < \infty$ and that $\mathbb{E}Y(s) = 1$. The former is straightforward whereas the second one while be proved by showing that it is a martingale. First of all $Y(s)$ is clearly integrable and moreover, considering a filtration $\mathcal{F}_s$, as a function of $\mathcal{F}_s$-measurable random

variables $Y(s)$ is also $\mathcal{F}_s$-measurable. For all $0 \leq s < t < \infty$ we have

$$\begin{aligned}
\mathbb{E}[\exp\{W(t) - t/2\}|\mathcal{F}_s] &= \exp\{W(s) - s/2\}\mathbb{E}\left[\frac{\exp\{W(t) - t/2\}}{\exp\{W(s) - s/2\}}\middle|\mathcal{F}_s\right] \\
&= \exp\{W(s) - s/2\}\mathbb{E}\left[\exp\{W(t) - W(s) - (t-s)/2\}\middle|\mathcal{F}_s\right] \\
&= \exp\{W(s) - s/2\}\mathbb{E}\left[\exp\{W(t) - W(s) - (t-s)/2\}\right] \\
&= \exp\{W(s) - s/2\}\int_{\mathbb{R}} \exp\left\{x + \frac{s-t}{2}\right\}\frac{1}{2\pi(s-t)}\exp\left\{-\frac{x^2}{2(s-t)}\right\}\mathrm{d}x \\
&= \exp\{W(s) - s/2\}\int_{\mathbb{R}} \frac{1}{2\pi(s-t)}\exp\left\{-\frac{(x-s+t)^2}{2(s-t)}\right\}\mathrm{d}x \\
&= \exp\{W(s) - s/2\}.
\end{aligned}$$

Thus $Y(s)$ is a martingale and so $\forall s$, $\mathbb{E}Y(s) = \mathbb{E}Y(0) = \mathbb{E}[\exp\{W(0) - 0/2\}] = 1$ which proves that the assumption $\mathbb{E}[Y(s)] = 1$. Thus conditions of Theorem 2.5 are met and the Brown-Resnick from Section 2.6.3 is appropriately defined as the limit of i.i.d. Ornstein-Ulhenbeck rocesses.

Appendix C

Proof of the expressions of the marginal distributions for the Brown-Resnick model

C.1 Proof of Theorem 3.5

Let $\eta'(s)$ be the process defined in (3.1) and consider firstly the case $0 \leq s_1 \leq s_2 \leq s_3$. According to Lemma 2.7 we have,

$$\begin{aligned}
&-\log \Pr(\eta'(s_1) \leq e^{x_1}, \eta'(s_2) \leq e^{x_2}, \eta'(s_3) \leq e^{x_3}) \\
&\quad = \mathbb{E}\left[\max\left(e^{\epsilon(\beta s_1)-\beta s_1/2-x_1}, e^{\epsilon(\beta s_2)-\beta s_2/2-x_2}, e^{\epsilon(\beta s_3)-\beta s_3/2-x_3}\right)\right] \\
&\quad = \mathbb{E}\left[e^{\epsilon(\beta s_1)-\beta s_1/2}\exp\left\{\max\left(-x_1, \epsilon_{2,1}-\tfrac{1}{2}a_{2,1}^2-x_2, \epsilon_{3,1}-\tfrac{1}{2}a_{3,1}^2-x_3\right)\right\}\right] \\
&\quad = \mathbb{E}\left[\exp\left\{\max\left(-x_1, \epsilon_{2,1}-\tfrac{1}{2}a_{2,1}^2-x_2, \epsilon_{3,1}-\tfrac{1}{2}a_{3,1}^2-x_3\right)\right\}\right]
\end{aligned}$$

where $\epsilon_{i,j} = \epsilon(\beta s_i) - \epsilon(\beta s_j)$. Note that the last two equalities are consequences of the independency of the exponential terms and the fact that the Brown-Resnick process is a martingale. By definition of Brownian motions, we know that $\epsilon_{2,1} \sim \mathcal{N}\left(0, a_{2,1}^2\right)$ and $\epsilon_{3,2} \sim \mathcal{N}\left(0, a_{3,2}^2\right)$ are independent of each other. Noting that we have the relationship $\epsilon_{3,1} = \epsilon_{3,2} + \epsilon_{2,1}$ then it is easy to obtain the joint densities

$$\begin{aligned}
f_{\epsilon_{2,1},\epsilon_{3,1}}(u,v) &= c\exp\left\{-\frac{1}{2}\left(\left(\frac{u}{a_{2,1}}\right)^2 + \left(\frac{v-u}{a_{3,2}}\right)^2\right)\right\} \\
f_{\epsilon_{2,1},\epsilon_{3,2}}(u,v) &= c\exp\left\{-\frac{1}{2}\left(\left(\frac{u}{a_{2,1}}\right)^2 + \left(\frac{v}{a_{3,2}}\right)^2\right)\right\} \\
f_{\epsilon_{3,1},\epsilon_{3,2}}(u,v) &= c\exp\left\{-\frac{1}{2}\left(\left(\frac{u-v}{a_{2,1}}\right)^2 + \left(\frac{v}{a_{3,2}}\right)^2\right)\right\}
\end{aligned}$$

where $c = (2\pi a_{3,2}a_{2,1})^{-1}$.

If we set $Y = \max\left(\exp\{-x_1\}, \exp\left\{\epsilon_{2,1} - \frac{1}{2}a_{2,1}^2 - x_2\right\}, \exp\left\{\epsilon_{3,1} - \frac{1}{2}a_{3,1}^2 - x_3\right\}\right)$ then the probability mass function (pmf) of Y is given by

$$\begin{aligned}
\Pr\left(Y = \exp\left\{-x_1\right\}\right) &= \Pr\left(\epsilon_{2,1} \leq x_2 - x_1 + \tfrac{1}{2}a_{2,1}^2, \epsilon_{3,1} \leq x_3 - x_1 + \tfrac{1}{2}a_{3,1}^2\right) \\
\Pr\left(Y = \exp\left\{\epsilon_{2,1} - \tfrac{1}{2}a_{2,1}^2 - x_2\right\}\right) &= \Pr\left(\epsilon_{2,1} \geq x_2 - x_1 + \tfrac{1}{2}a_{2,1}^2, \epsilon_{3,2} \leq x_3 - x_2 + \tfrac{1}{2}a_{3,2}^2\right) \\
\Pr\left(Y = \exp\left\{\epsilon_{3,1} - \tfrac{1}{2}a_{3,1}^2 - x_3\right\}\right) &= \Pr\left(\epsilon_{3,1} \geq x_3 - x_1 + \tfrac{1}{2}a_{3,1}^2, \epsilon_{3,2} \geq x_2 - x_3 + \tfrac{1}{2}a_{2,1}^2\right)
\end{aligned}$$

Using Lemma 3.3, its expected value is then evaluated as follows

$$\begin{aligned}
\mathbb{E}[Y] &= e^{-x_1}\int_{-\infty}^{x_3-x_1+a_{3,1}^2/2}\int_{-\infty}^{x_2-x_1+a_{2,1}^2/2} f_{\epsilon_{2,1},\epsilon_{3,1}}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad + e^{-x_2}\int_{-\infty}^{x_3-x_2+a_{3,1}^2/2}\int_{x_2-x_1+a_{2,1}^2/2}^{\infty} \exp\left\{u - \tfrac{1}{2}a_{2,1}^2\right\} f_{\epsilon_{2,1},\epsilon_{3,2}}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad + e^{-x_3}\int_{x_3-x_2+a_{2,1}^2/2}^{\infty}\int_{x_3-x_1+a_{3,1}^2/2}^{\infty} \exp\left\{u - \tfrac{1}{2}a_{3,1}^2\right\} f_{\epsilon_{3,1},\epsilon_{3,2}}(u,v)\mathrm{d}u\mathrm{d}v \\
&= e^{-x_1}\int_{-\infty}^{x_3-x_1+a_{3,1}^2/2}\int_{-\infty}^{x_2-x_1+a_{2,1}^2/2} f_{\epsilon_{2,1},\epsilon_{3,1}}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad + e^{-x_2}\int_{-\infty}^{x_3-x_2+a_{3,1}^2/2}\int_{-\infty}^{x_1-x_2-a_{2,1}^2/2} c\exp\left\{-\frac{1}{2}\left(\left(\frac{u+a_{2,1}^2}{a_{2,1}}\right)^2 + \left(\frac{v}{a_{3,2}}\right)^2\right)\right\}\mathrm{d}u\mathrm{d}v \\
&\quad + e^{-x_3}\int_{-\infty}^{x_2-x_3-a_{2,1}^2/2}\int_{-\infty}^{x_1-x_3-a_{3,1}^2/2} c\exp\left\{-\frac{1}{2}\left(\left(\frac{u-v+a_{2,1}^2}{a_{2,1}}\right)^2 + \left(\frac{v+a_{3,2}^2}{a_{3,2}}\right)^2\right)\right\}\mathrm{d}u\mathrm{d}v \\
&= e^{-x_1}\int_{-\infty}^{x_2-x_1+a_{2,1}^2/2}(2\pi a_{2,1})^{-1}\phi(u/a_{2,1})\Phi\left(\frac{x_3-x_1+a_{3,1}^2/2-u}{a_{3,2}}\right)\mathrm{d}u \\
&\quad + e^{-x_2}\Phi\left((x_3-x_2)a_{3,2}^{-1} + \tfrac{1}{2}a_{3,2}\right)\Phi\left((x_1-x_2)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right) \\
&\quad + e^{-x_3}\int_{-\infty}^{x_2-x_3-a_{2,1}^2/2}(2\pi a_{3,2})^{-1}\phi(v/a_{3,2}+a_{3,2}^2)\Phi\left(\frac{x_1-x_3-\frac{1}{2}a_{3,1}^2-v+a_{2,1}^2}{a_{2,1}}\right)\mathrm{d}v \\
&= e^{-x_1}\Phi_2\left(\left\{(x_3-x_1)a_{3,1}^{-1} + \tfrac{1}{2}a_{3,1}, (x_2-x_1)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right\}^\top; a_{2,1}/a_{3,1}\right) \\
&\quad + e^{-x_2}\Phi_2\left(\left\{(x_3-x_2)a_{3,2}^{-1} + \tfrac{1}{2}a_{3,2}, (x_1-x_2)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right\}^\top; 0\right) \\
&\quad + e^{-x_3}\Phi_2\left(\left\{(x_2-x_3)a_{3,2}^{-1} + \tfrac{1}{2}a_{3,2}, (x_1-x_3)a_{3,1}^{-1} + \tfrac{1}{2}a_{3,1}\right\}^\top; a_{3,2}/a_{3,1}\right),
\end{aligned} \tag{C.1}$$

which completes the proof for $0 \leq s_1 \leq s_2 \leq s_3$. The reasoning is similar for any s_1, s_2 and $s_3 \in \mathbb{R}$ as soon as they are either all positive or all negative.

Now let us consider the case where one of the s_i, $i = 1, 2, 3$ has a different sign as the other

two. For example we will consider the case where $s_1 \leq 0 \leq s_2 \leq s_3$.

$$
\begin{aligned}
&-\log \Pr(\eta(s_1) \leq e^{x_1}, \eta(s_2) \leq e^{x_2}, \eta(s_3) \leq e^{x_3}) \\
&\quad = \mathbb{E}\left[\exp\left\{\max\left(\epsilon(\beta s_1) + \tfrac{1}{2}\beta s_1 - x_1, \epsilon(\beta s_2) - \tfrac{1}{2}\beta s_2 - x_2, \epsilon(\beta s_3) - \tfrac{1}{2}\beta s_3 - x_3\right)\right\}\right] \\
&\quad = \int_0^\infty k \mathrm{dPr}\left(\exp\left\{\max\left(\epsilon(\beta s_1) + \tfrac{1}{2}\beta s_1 - x_1, \epsilon(\beta s_2) - \tfrac{1}{2}\beta s_2 - x_2, \epsilon(\beta s_3) - \tfrac{1}{2}\beta s_3 - x_3\right)\right\} \leq k\right).
\end{aligned} \tag{C.2}
$$

Moreover the condition $s_1 \leq 0 \leq s_2 \leq s_3$ implies that $\epsilon(\beta s_1)$ is independent of both $\epsilon(\beta s_2)$ and $\epsilon(\beta s_3)$ and thus

$$
\begin{aligned}
&\Pr(\exp\{\max(\epsilon(\beta s_1) + \tfrac{1}{2}\beta s_1 - x_1, \epsilon(\beta s_2) - \tfrac{1}{2}\beta s_2 - x_2, \epsilon(\beta s_3) - \tfrac{1}{2}\beta s_3 - x_3)\} \leq k) \\
&= \Pr\left(\epsilon(\beta s_1) - \tfrac{1}{2}\beta|s_1| - x_1 \leq \log k\right) \\
&\qquad\qquad \times \Pr\left(\epsilon(\beta s_2) - \tfrac{1}{2}\beta s_2 - x_2 \leq \log k, \epsilon(\beta s_3) - \tfrac{1}{2}\beta s_3 - x_3 \leq \log k\right)
\end{aligned}
$$

It is easy to see that

$$
\Pr(\epsilon(\beta s_1) - \beta|s_1|/2 - x_1 \leq \log k) = \Phi\left((\log k + x_1 + \tfrac{1}{2}\beta|s_1|)/\sqrt{\beta|s_1|}\right) := F_1(k).
$$

Also, because the pdf of $(\epsilon(\beta s_2), \epsilon(\beta s_3) - \epsilon(\beta s_2))$ is given by

$$
f_{\epsilon(\beta s_2), \epsilon(\beta s_3) - \epsilon(\beta s_2)}(u, v) = \left(2\pi\beta\sqrt{s_2(s_3 - s_2)}\right)^{-1} \exp\left\{-\frac{1}{2}\left(\frac{u^2}{\beta s_2} + \frac{v^2}{\beta(s_3 - s_2)}\right)\right\},
$$

we then get

$$
\begin{aligned}
&\Pr(\epsilon(\beta s_2) - \beta s_2/2 - x_2 \leq \log k, \epsilon(\beta s_3) - \beta s_3/2 - x_3 \leq \log k) \\
&\quad = \Phi_2\left(\left\{(\log k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}, (\log k + x_3 + \beta s_3/2)/\sqrt{\beta s_3}\right\}^\top, -\sqrt{s_2/s_3}\right) := F_2(k, k)
\end{aligned}
$$

Additionally, differentiating the product $F_1(k)F_2(k,k)$ w.r.t k, we obtain

$$
\mathrm{d}F_1(k)F_2(k,k) = F_1'(k)F_2(k,k) + F_1(k)\frac{\partial}{\partial k}F_2(k,k), \tag{C.3}
$$

where

$$
F_1'(k) = \left(k\sqrt{\beta|s_1|}\right)^{-1} \phi\left((\log k + x_1 + \beta|s_1|/2)/\sqrt{\beta|s_1|}\right)
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial k}F_2(k,k) = &\frac{1}{k\sqrt{\beta s_2}} \int_{-\infty}^{(\log k + x_3 + \beta s_3/2)/\sqrt{\beta s_3}} \phi_2\left(\left\{u, (\log k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right) \mathrm{d}u \\
&+ \frac{1}{k\sqrt{\beta s_3}} \int_{-\infty}^{(\log k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}} \phi_2\left(\left\{(\log k + x_3 + \beta s_3/2)/\sqrt{\beta s_3}, v\right\}^\top; \sqrt{s_2/s_3}\right) \mathrm{d}v.
\end{aligned}
$$

Thus, plugging (C.3) into (C.2) leads to,

$$\begin{aligned}
&\int_0^\infty \frac{1}{\sqrt{\beta|s_1|}}\phi\left((\log k + x_1 + \tfrac{1}{2}\beta|s_1|)/\sqrt{\beta|s_1|}\right) \\
&\quad\times \int_0^k\int_0^k \phi_2\left(\left\{(\log u + x_3 + \tfrac{1}{2}\beta s_3)/\sqrt{\beta s_3}, (\log v + x_2 + \tfrac{1}{2}\beta s_2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)\frac{\mathrm{d}u\mathrm{d}v\mathrm{d}k}{uv\beta\sqrt{s_2 s_3}} \\
&+\int_0^\infty \Phi\left(\left(\log k + x_1 + \tfrac{1}{2}\beta|s_1|\right)/\sqrt{\beta|s_1|}\right) \\
&\quad\times \int_0^k \phi_2\left(\left\{(\log u + x_3 + \tfrac{1}{2}\beta s_3)/\sqrt{\beta s_3}, (\log k + x_2 + \tfrac{1}{2}\beta s_2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)\frac{\mathrm{d}u\mathrm{d}k}{u\beta\sqrt{s_2 s_3}} \\
&+\int_0^\infty \Phi\left(\left(\log k + x_1 + \tfrac{1}{2}\beta|s_1|\right)/\sqrt{\beta|s_1|}\right) \\
&\quad\times \int_0^k \phi_2\left(\left\{(\log k + x_3 + \tfrac{1}{2}\beta s_3)/\sqrt{\beta s_3}, (\log v + x_2 + \tfrac{1}{2}\beta s_2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)\frac{\mathrm{d}v\mathrm{d}k}{v\beta\sqrt{s_2 s_3}}.
\end{aligned} \tag{C.4}$$

We can now calculate separately each of the three components above. Applying Fubini's theorem, Lemmas 3.3 and 3.4 and the change of variables $u' = e^u$, $v' = e^v$, the first component of (C.4) can be re-written as

$$\begin{aligned}
&\int_0^\infty \int_{-\infty}^{\log k}\int_{-\infty}^{\log k} (\beta|s1|)^{-1/2}\phi\left((\log k + x_1 + \beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi\left((v + x_2 + \beta s_2/2)/\sqrt{\beta s_2}\right) \\
&\quad\times\phi\left((u - v + x_3 - x_2 + (s_3 - s_2)/2)/\sqrt{\beta(s_3 - s_2)}\right)\frac{\mathrm{d}u\mathrm{d}v\mathrm{d}k}{\beta\sqrt{s_2(s_3 - s_2)}} \\
=&\int_0^\infty \int_{-\infty}^{(\log k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}} (\beta|s_1|)^{-1/2}\phi\left((\log k + x_1 + \beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi(v) \\
&\quad\times\Phi\left(\left(\log k - v\sqrt{\beta s_2} + x_3 + \beta s_3/2\right)/\sqrt{\beta(s_3 - s_2)}\right)\mathrm{d}v\mathrm{d}k \\
=&\int_{-\infty}^\infty (\beta|s_1|)^{-1/2}\phi\left((k + x_1 + \beta|s_1|/2)/\sqrt{\beta|s_1|}\right) \\
&\quad\times\Phi_2\left(\left\{(k + x_3 + \beta s_3/2)/\sqrt{\beta s_3}, (k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)e^k\mathrm{d}k \\
=&e^{-x_1}\int_{-\infty}^\infty (\beta|s_1|)^{-1/2}\phi\left((k + x_1 - \beta|s_1|/2)/\sqrt{\beta|s_1|}\right) \\
&\quad\times\Phi_2\left(\left\{(k + x_3 + \beta s_3/2)/\sqrt{\beta s_3}, (k + x_2 + \beta s_2/2)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)\mathrm{d}k \\
=&e^{-x_1}\int_{-\infty}^\infty \phi(k)\Phi_2\left(\left\{\left(k\sqrt{|s_1|} + x_3 - x_1 + \beta(s_3 + |s_1|)/2\right)/\sqrt{\beta s_3},\right.\right. \\
&\qquad\left.\left.\left(k\sqrt{|s_1|} + x_2 - x_1 + \beta(s_2 + |s_1|)/2\right)/\sqrt{\beta s_2}\right\}^\top; \sqrt{s_2/s_3}\right)\mathrm{d}k \\
=&\Phi_2\left(\left\{(x_2 - x_1)a_{3,1}^{-1} + \tfrac{1}{2}a_{3,1}, (y - x)a_{2,1}^{-1} + \tfrac{1}{2}a_{2,1}\right\}^\top; a_{2,1}/a_{3,1}\right).
\end{aligned} \tag{C.5}$$

Similarly as above, the second term of (C.4) can be firstly transformed using Fubini's theorem,

Lemmas 3.3 and 3.4 and the change of variable $v' = e^v$ leading to

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{k} \Phi\left((k+x_1+\beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi\left((k+y_2+\beta s_2/2)/\sqrt{\beta s_2}\right)\\
&\quad\times\phi\left((v-k+x_3-x_2+\beta(s_3-s_2)/2)/\sqrt{\beta(s_3-s_2)}\right)e^v\frac{\mathrm{d}v\mathrm{d}k}{\beta\sqrt{s_2(s_3-s_2)}}\\
=&e^{-x_2}\int_{-\infty}^{\infty}\int_{-\infty}^{k} \Phi\left((k+x_1+\beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi\left((k+y_2-\beta s_2/2)/\sqrt{\beta s_2}\right)\\
&\quad\times\phi\left((v-k+x_3-x_2+\beta(s_3-s_2)/2)/\sqrt{\beta(s_3-s_2)}\right)\frac{\mathrm{d}v\mathrm{d}k}{\beta\sqrt{s_2(s_3-s_2)}}\\
=&e^{-x_2}\Phi\left((x_3-x_2+(s_3-s_2)/2)\sqrt{s_3-s_2}\right)\\
&\quad\times\int_{-\infty}^{\infty}\Phi\left((k+x_1+\beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi((v+x_2-s_2/2)\sqrt{s_2})\frac{\mathrm{d}k}{\sqrt{s_2}}\\
=&e^{-x_2}\Phi((x_3-x_2+(s_3-s_2)/2)\sqrt{s_3-s_2})\\
&\quad\times\Phi_2\left(\left\{\infty,(x_1-x_2+(s_2+|s_1|))/\sqrt{s_2+|s_1|}\right\}^\top;-\sqrt{s_2/(s_2+|s_1|)}\right)\\
=&e^{-x_2}\Phi\left((x_3-x_2)a_{3,2}^{-1}+\tfrac{1}{2}a_{3,2}\right)\Phi\left((x_1-x_2)a_{2,1}^{-1}+\tfrac{1}{2}a_{2,1}\right).
\end{aligned}$$

Finally, following the same method, the third and last term of (C.4) can be reduced to

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\log k} \Phi\left((k+x_1+\beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi\left((k+x_3+\beta s_3/2)/\sqrt{\beta s_3}\right)\\
&\quad\times\phi\left((\beta s_3(v+x_2)-\beta s_2(k+x_3))/\sqrt{\beta^3 s_2 s_3(s_3-s_2)}\right)e^k\frac{\mathrm{d}v\mathrm{d}k}{\beta\sqrt{s_2(s_3-s_2)}}\\
=&e^{-x_3}\int_{-\infty}^{\infty}\int_{-\infty}^{\log k} \Phi\left((k+x_1+\beta|s_1|/2)/\sqrt{\beta|s_1|}\right)\phi\left((k+x_3-\beta s_3/2)/\sqrt{\beta s_3}\right)\\
&\quad\times\phi\left((\beta s_3(v+x_2)-\beta s_2(k+x_3))/\sqrt{\beta^3 s_2 s_3(s_3-s_2)}\right)\frac{\mathrm{d}v\mathrm{d}k}{\beta\sqrt{s_2(s_3-s_2)}}\\
=&e^{-x_3}\int_{-\infty}^{\infty}\frac{1}{\sqrt{\beta s_3}}\phi\left((k+x_3-\beta s_3/2)/\sqrt{\beta s_3}\right)\\
&\quad\times\Phi_2\left(\left\{(\beta(s_3-s_2)\log k+\beta s_3x_2-\beta s_2x_3)/\sqrt{\beta^3 s_2 s_3(s_3-s_2)},\left(k+x_1+\tfrac{1}{2}\beta|s_1|\right)/\sqrt{\beta|s_1|}\right\}^\top;0\right)\mathrm{d}k\\
=&e^{-x_3}\Phi_2\left(\left\{(x_2-x_3)a_{3,2}^{-1}+\tfrac{1}{2}a_{3,2},(x_1-x_3)a_{3,1}^{-1}+\tfrac{1}{2}a_{3,1}\right\}^\top;a_{3,2}/a_{3,1}\right).
\end{aligned}$$

This completes the proof for the case $s_1 \le 0 \le s_2 \le s_3$. The reasoning is similar whenever one of the $s_i, i = 1, 2, 3$ has a different sign than the other two.

C.2 Proof of Lemma 3.6

Assume $s_1 \le \ldots \le s_p$ and let $\epsilon(s_i)$ denote a Brownian motion at time $s_i, i = 1, \ldots, p$. From the properties of Brownian motions we know that, for $1 \le n \le p$,

$$\epsilon = (\epsilon_{n,1}, \ldots, \epsilon_{n,n-1}, \epsilon_{n+1,n}, \ldots, \epsilon_{p,n}) \sim \mathcal{N}_{p-1}(0, \Sigma),$$

where $\epsilon_{i,j} = \epsilon(s_i) - \epsilon(s_j)$ for $i, j = 1, \ldots, p, i \neq j$. Brownian motions have the property of having independent increments, hence for $1 \leq i, j, n, \leq p$, if $i < j < n$ then,

$$\begin{aligned}
\mathrm{Cov}(\epsilon_{n,i}, \epsilon_{n,j}) &= \mathrm{Cov}(\epsilon(s_n) - \epsilon(s_j) + \epsilon(s_j) - \epsilon(s_i), \epsilon(s_n) - \epsilon(s_j)) \\
&= \mathrm{Cov}(\epsilon(s_n) - \epsilon(s_j), \epsilon(s_n) - \epsilon(s_j)) + \mathrm{Cov}(\epsilon(s_j) - \epsilon(s_i), \epsilon(s_n) - \epsilon(s_j)) \\
&= \mathrm{Var}(\epsilon(s_n) - \epsilon(s_j)) \\
&= s_n - s_j = s_{n,j},
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathrm{Cov}(\epsilon_{n,i}, \epsilon_{n,j}) = s_{n,i}, \quad \text{if } j < i < n, \qquad \mathrm{Cov}(\epsilon_{n,i}, \epsilon_{n,j}) = s_{i,n}, \quad \text{if } n < i < j, \\
&\mathrm{Cov}(\epsilon_{n,i}, \epsilon_{n,j}) = s_{j,n}, \quad \text{if } n < j < i, \qquad \mathrm{Cov}(\epsilon_{n,i}, \epsilon_{j,n}) = 0, \quad \text{if } i < n < j \text{ or } j < n < i.
\end{aligned}$$

Thus the covariance matrix of the joint distribution is

$$\Sigma = \left[\begin{array}{cccc:ccc}
s_{n,1} & s_{n,2} & \cdots & s_{n,n-1} & 0 & \cdots & 0 & 0 \\
s_{n,2} & s_{n,2} & \cdots & s_{n,n-1} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\
s_{n,n-1} & s_{n,n-1} & \cdots & s_{n,n-1} & 0 & \cdots & 0 & 0 \\
\hdashline
0 & 0 & \cdots & 0 & s_{n+1,n} & \cdots & s_{n+1,n} & s_{n+1,n} \\
\vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & s_{n+1,n} & \cdots & s_{p-1,n} & s_{p-1,n} \\
0 & 0 & \cdots & 0 & s_{n+1,n} & \cdots & s_{p-1,n} & s_{p,n}
\end{array}\right] = \left[\begin{array}{c:c} \Sigma_1 & 0 \\ \hdashline 0 & \Sigma_2 \end{array}\right]$$

where $\Sigma_{1,1}$ is a $(n-1)\times(n-1)$ matrix and $\Sigma_{2,2}$ a $(p-n)\times(p-n)$ matrix. Consequently, ϵ can be expressed as the product of two normally distributed random processes $\epsilon_1 = (\epsilon_{n,1}, \ldots, \epsilon_{n,n-1})$ and $\epsilon_2 = (\epsilon_{n+1,n}, \ldots, \epsilon_{p,n})$, respectively with densities $\phi_{n-1}(u; \Sigma_1), u \in \mathbb{R}^{n-1}$ and $\phi_{p-n}(v; \Sigma_2), v \in \mathbb{R}^{p-n}$.

Thus the expected value of Z is

$$\mathbb{E}[Z] = e^{-x_n} \mathrm{I}_1 \times \mathrm{I}_2 \tag{C.6}$$

where

$$\mathrm{I}_1 = \int_{x_{n,n-1}+\frac{1}{2}s_{n,n-1}}^{\infty} \cdots \int_{x_{n,1}+\frac{1}{2}s_{n,1}}^{\infty} e^{u_1 - \frac{s_{d,1}}{2}} \phi_{n-1}(u; \Sigma_1) \mathrm{d}u,$$

and

$$\mathrm{I}_2 = \int_{-\infty}^{x_{p,n}+\frac{1}{2}s_{p,n}} \cdots \int_{-\infty}^{x_{n+1,n}+\frac{1}{2}s_{n+1,n}} \phi_{p-n}(v; \Sigma_2) \mathrm{d}v.$$

Before going any further two results, also obtained by induction, need to be stated. First, we

prove that the following hypothesis:

$$\mathrm{H}'_n : \text{If } \Sigma_n = \begin{bmatrix} s_{n,1} & s_{n,2} & \cdots & s_{n,n-1} \\ s_{n,2} & s_{n,2} & \cdots & s_{n,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,n-1} & s_{n,n-1} & \cdots & s_{n,n-1} \end{bmatrix} \text{ then its inverse is}$$

$$\Sigma_n^{-1} = \begin{bmatrix} \frac{1}{s_{2,1}} & \frac{-1}{s_{2,1}} & 0 & \cdots & 0 & 0 & 0 \\ \frac{-1}{s_{2,1}} & \frac{s_{3,1}}{s_{3,2}s_{2,1}} & \frac{-1}{s_{3,2}} & \cdots & 0 & 0 & 0 \\ 0 & \frac{-1}{s_{3,2}} & \frac{s_{4,2}}{s_{4,3}s_{3,2}} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{s_{n-2,n-4}}{s_{n-2,n-3}s_{n-3,n-4}} & \frac{-1}{s_{n-2,n-3}} & 0 \\ 0 & 0 & 0 & \cdots & \frac{-1}{s_{n-2,n-3}} & \frac{s_{n-1,n-3}}{s_{n-1,n-2}s_{n-2,n-3}} & \frac{-1}{s_{n-1,n-2}} \\ 0 & 0 & 0 & \cdots & 0 & \frac{-1}{s_{n-1,n-2}} & \frac{s_{n,n-2}}{s_{n,n-1}s_{n-1,n-2}} \end{bmatrix} \tag{C.7}$$

is true for all $n \geq 2$.

Let $n = 2$, then $\mathrm{H}'_2 : \Sigma_2 = s_{2,1}$ and $\Sigma_3^{-1} = s_{2,1}^{-1}$.

Let $n = 3$, then $\mathrm{H}'_3 : \Sigma_3 = \begin{bmatrix} s_{3,1} & s_{3,2} \\ s_{3,2} & s_{3,2} \end{bmatrix}$ and $\Sigma_3^{-1} = \frac{1}{s_{3,2}s_{2,1}} \begin{bmatrix} s_{3,2} & -s_{3,2} \\ -s_{3,2} & s_{3,1} \end{bmatrix} = \begin{bmatrix} \frac{1}{s_{2,1}} & \frac{-1}{s_{2,1}} \\ \frac{-1}{s_{2,1}} & \frac{s_{3,1}}{s_{3,2}s_{2,1}} \end{bmatrix}$.

Assume H'_n true for all k, $2 \leq k \leq n$ we consider the case $k+1$. Here

$$\Sigma_{k+1} = \left[\begin{array}{cccc:c} s_{k+1,1} & s_{k+1,2} & \cdots & s_{k,k-1} & s_{k+1,k} \\ s_{k+1,2} & s_{k+1,2} & \cdots & s_{k,k-1} & s_{k+1,k} \\ \vdots & \vdots & \ddots & \vdots & \\ s_{k,k-1} & s_{k,k-1} & \cdots & s_{k,k-1} & s_{k,k-1} \\ \hdashline s_{k+1,k} & s_{k+1,k} & \cdots & s_{k+1,k} & s_{k+1,k} \end{array}\right] = \left[\begin{array}{c:c} s_{k+1,k}J_{k-1} + \Sigma_k & s_{k+1,k}\mathbb{1}_{k-1} \\ \hdashline s_{k+1,k}\mathbb{1}_{k-1}^\top & s_{k+1,k} \end{array}\right]$$

where J_{k-1} is a $(k-1) \times (k-1)$ unit matrix. Using the block representation of Σ_{k+1} and applying Schur's decomposition leads to

$$\begin{aligned} \Sigma_{k+1}^{-1} &= \begin{bmatrix} J_{k-1} & 0 \\ -\mathbb{1}_{k-1}^\top & 1 \end{bmatrix} \begin{bmatrix} \Sigma_k^{-1} & 0 \\ 0 & \frac{1}{s_{k+1,k}} \end{bmatrix} \begin{bmatrix} J_{k-1} & -\mathbb{1}_{k-1} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_k^{-1} & 0 \\ -\mathbb{1}_{k-1}^\top \Sigma_k^{-1} & \frac{1}{s_{k+1,k}} \end{bmatrix} \begin{bmatrix} J_{k-1} & -\mathbb{1}_{k-1} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_k^{-1} & -\Sigma_k^{-1}\mathbb{1}_{k-1} \\ -\mathbb{1}_{k-1}^\top \Sigma_k^{-1} & \mathbb{1}_{k-1}^\top \Sigma_k^{-1}\mathbb{1}_{k-1} + \frac{1}{s_{k+1,k}} \end{bmatrix}. \end{aligned}$$

Using the induction hypothesis H'_k we find

$$-\mathbb{1}_{k-1}^\top \Sigma_k^{k-1} = \left(0, \dots, 0, -\frac{1}{s_{k,k-1}}\right)$$

a row vector of length $(k-1)$ and

$$\mathbb{1}_{k-1}^\top \Sigma_k^{-1} \mathbb{1}_{k-1} + \frac{1}{s_{k+1,k}} = \frac{1}{s_{k,k-1}} + \frac{1}{s_{k+1,k}} = \frac{s_{k+1,k-1}}{s_{k+1,k} s_{k,k-1}}.$$

Consequently we finally obtain

$$\Sigma_{k+1}^{-1} = \left[\begin{array}{ccc:c} & & & 0 \\ & \Sigma_k^{-1} & & \vdots \\ & & & 0 \\ & & & -\frac{1}{s_{k,k-1}} \\ \hdashline 0 & \cdots & 0 & \frac{-1}{s_{k,k-1}} \quad \frac{s_{k+1,k-1}}{s_{k+1,k} s_{k,k-1}} \end{array}\right]$$

which proves H'_n true for all $n \geq 2$. The second hypothesis to verify is the following,

$$\mathrm{H}_n : \left\{ \begin{array}{l} \mathrm{Det}(\Sigma_1) = \prod_{i=1}^{n-1} s_{i+1,i} \\ t_{n,1} - 2u_1 + u_n^\top \Sigma_1^{-1} u_n = \sum_{i=1}^{n-2} \left(\frac{u_i - u_{i+1} - s_{i+1,i}}{s_{i+1,i}^{1/2}} \right)^2 + \left(\frac{u_{n-1} - s_{n,n-1}}{s_{n,n-1}^{1/2}} \right)^2 \end{array} \right. \tag{C.8}$$

where $u_n = (u_1, \dots, u_{n-1})$. and we can now do another induction to prove H_n.

Let $n = 2$, then

$$\mathrm{H}_2 : \left\{ \begin{array}{l} \mathrm{Det}(s_{2,1}) = s_{2,1} \\ s_{2,1} - 2u_1 + u_1^2 s_{2,1}^{-1} = \left(\frac{u_1 - s_{2,1}}{s_{2,1}^{1/2}} \right)^2 \end{array} \right. .$$

Let $n = 3$, then

$$\mathrm{H}_3 : \left\{ \begin{array}{l} \mathrm{Det} \begin{pmatrix} s_{3,1} & s_{3,2} \\ s_{3,2} & s_{3,2} \end{pmatrix} = s_{3,2} s_{2,1} \\ s_{3,1} - 2u_1 + (u_1, u_2)^\top \begin{pmatrix} \frac{1}{s_{2,1}} & -\frac{1}{s_{2,1}} \\ -\frac{1}{s_{2,1}} & \frac{s_{3,1}}{s_{3,2} s_{2,1}} \end{pmatrix} (u_1, u_2) = \left(\frac{u_1 - u_2 - t_{2,1}}{t_{2,1}^{1/2}} \right)^2 + \left(\frac{u_2 - t_{3,2}}{t_{3,2}^{1/2}} \right)^2 \end{array} \right. .$$

Assume H_n true for all k, $2 \leq k \leq n$ and consider the case $k+1$.

$$\mathrm{H}_{k+1} : \left\{ \begin{array}{l} \mathrm{Det}(\Sigma_{k+1}) = \prod_{i=1}^{k} s_{i+1,i} \\ s_{k+1,1} - 2u_1 + u_{k+1}^\top \Sigma_{k+1}^{-1} u_{k+1} = \sum_{i=1}^{k-1} \left(\frac{u_i - u_{i+1} - s_{i+1,i}}{s_{i+1,i}^{1/2}} \right)^2 + \left(\frac{u_k - s_{k+1,k}}{s_{k+1,k}^{1/2}} \right)^2 \end{array} \right. .$$

If we focus on the first element of H_{k+1}, using once again Schur's decomposition we have

$$\Sigma_{k+1}^{-1} = \begin{bmatrix} J_{k-1} & 0 \\ -\mathbb{1}_{k-1}^\top & 1 \end{bmatrix} \begin{bmatrix} \Sigma_k^{-1} & 0 \\ 0 & s_{k+1,k}^{-1} \end{bmatrix} \begin{bmatrix} J_{k-1} & -\mathbb{1}_{k-1} \\ 0 & 1 \end{bmatrix}$$

which leads to,

$$\begin{aligned} \mathrm{Det}\left(\Sigma_{k+1}^{-1}\right) &= \mathrm{Det}\begin{pmatrix} J_{k-1} & 0 \\ -\mathbb{1}_{k-1}^\top & 1 \end{pmatrix} \times \mathrm{Det}\begin{pmatrix} \Sigma_k^{-1} & 0 \\ 0 & s_{k+1,k}^{-1} \end{pmatrix} \times \mathrm{Det}\begin{pmatrix} J_{k-1} & -\mathbb{1}_{k-1} \\ 0 & 1 \end{pmatrix} \\ &= 1 \times \mathrm{Det}\left(\Sigma_k^{-1}\right) \times s_{k+1,k}^{-1} \times 1 \\ &= \left(\mathrm{Det}(\Sigma_k) s_{k+1,k}\right)^{-1} \\ &= \left(s_{k+1,k} s_{k,k-1} \dots s_{2,1}\right)^{-1}, \end{aligned}$$

and consequently the assertion $\mathrm{Det}(\Sigma_{k+1}) = \prod_{i=1}^{k} s_{i+1,i}$ is proved.

If we look at the second element of H_{k+1}, applying the induction hypothesis gives

$$\begin{aligned} & s_{k+1,1} - 2u_1 + u_{k+1}^\top \Sigma_{k+1}^{-1} u_{k+1} \\ &\quad = s_{k+1,k} + s_{k,1} - 2u_1 + u_k^\top \Sigma_k^{-1} u_k - 2\frac{u_{k-1}u_k}{s_{k,k-1}} + u_k^2 \frac{s_{k+1,k-1}}{s_{k+1,k}s_{k,k-1}} \\ &\quad = \sum_{i=1}^{k-2} \left(\frac{u_i - u_{i+1} - s_{i+1,i}}{s_{i+1,i}^{1/2}} \right)^2 + \left(\frac{u_{k-1} - s_{k,k-1}}{s_{k,k-1}^{1/2}} \right)^2 - 2\frac{u_{k-1}u_k}{s_{k,k-1}} + \frac{u_k^2}{s_{k,k-1}} + \frac{u_k^2}{s_{k+1,k}} + s_{k+1,k} \\ &\quad = \sum_{i=1}^{k-1} \left(\frac{u_i - u_{i+1} - s_{i+1,i}}{s_{i+1,i}^{1/2}} \right)^2 + \left(\frac{u_k - s_{k+1,k}}{s_{k+1,k}^{1/2}} \right)^2 \end{aligned}$$

which proves that H_n is true for all $n \geq 2$.

Going back to the first term of (C.6), considering also the change of variable $u_i' = -u_i, i = 1, \dots, n-1$ we now have

$$\begin{aligned} \mathrm{I}_1 = \int_{-\infty}^{x_{n-1,n} - \frac{1}{2}s_{n,n-1}} \cdots \int_{-\infty}^{x_{1,n} - \frac{1}{2}s_{n,1}} & (2\pi)^{-\frac{(n-1)}{2}} \prod_{i=1}^{n-1} s_{i+1,i}^{-1/2} \\ & \times \exp\left\{ -\frac{1}{2} \left[\sum_{i=1}^{n-2} \left(\frac{u_i - u_{i+1} + s_{i+1,i}}{t_{i+1,i}^{1/2}} \right)^2 + \left(\frac{u_{n-1} + s_{n,n-1}}{s_{n,n-1}^{1/2}} \right)^2 \right] \right\} \mathrm{d}u_1 \cdots \mathrm{d}u_{n-1}. \quad \text{(C.9)} \end{aligned}$$

The substitution $v_i = \frac{u_i - u_{i+1} + s_{i+1,i}}{s_{i+1,i}^{1/2}}$ for $i = 1, \dots, n-2$ and $v_{n-1} = \frac{u_{n-1} + s_{n,n-1}}{s_{i+1,i}^{1/2}}$ gives

$$u_j = \sum_{i=1}^{n-j} \left\{ \left(v_{n-i} s_{n-i+1,n-i}^{1/2} \right) - s_{n-i+1,n-i} \right\}, \quad 1 \leq j \leq d-1,$$

and

$$\frac{\mathrm{d}v_1 \cdots \mathrm{d}v_{n-1}}{\mathrm{d}u_2 \cdots \mathrm{d}u_{n-1}} = \begin{vmatrix} s_{2,1}^{-1/2} & -s_{2,1}^{-1/2} & 0 & \cdots & 0 & 0 & 0 \\ -s_{2,1}^{-1/2} & s_{3,2}^{-1/2} & -s_{3,2}^{-1/2} & \cdots & 0 & 0 & 0 \\ 0 & -s_{3,2}^{-1/2} & s_{4,3}^{-1/2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & s_{n-2,n-3}^{-1/2} & -s_{n-2,n-3}^{-1/2} & 0 \\ 0 & 0 & 0 & \cdots & -s_{n-2,n-3}^{-1/2} & s_{n-1,n-2}^{-1/2} & -s_{n-1,n-2}^{-1/2} \\ 0 & 0 & 0 & \cdots & 0 & -s_{n-1,n-2}^{-1/2} & s_{n,n-1}^{-1/2} \end{vmatrix} = \prod_{i=1}^{n-1} s_{i+1,i}^{-1/2}.$$

Moreover the new bounds of integration are given by

$$-\infty \leq v_j \leq \left(x_{j,n} + \frac{s_{n,j}}{2} - \sum_{l=1}^{n-j-1} v_{n-l} s_{n-l+1,n-l}^{1/2} \right) s_{j+1,j}^{-1/2}, \quad 1 \leq j \leq n-2,$$

and

$$-\infty \leq \left(v_{n-1} \leq x_{n-1,n} + \tfrac{t_{n,n-1}}{2} \right) t_{d,d-1}^{-1/2}.$$

Thus following from (C.9),

$$\mathrm{I}_1 = \int_{-\infty}^{x_{n-1,n} s_{n,n-1}^{-1/2} + s_{n,n-1}^{1/2}/2} \int_{-\infty}^{(x_{n-2,n} + s_{n,n-2}/2 - v_{n-1} s_{n,n-1}) s_{n-1,n-2}^{-1/2}} \cdots$$
$$\int_{-\infty}^{(x_{1,n} + s_{n,1}/2 - \sum_{l=1}^{n-j-1} v_{n-l} s_{n-l+1,n-l}^{1/2}) s_{2,1}^{-1/2}} (2\pi)^{-\frac{(n-1)}{2}} \exp\left\{ -\frac{(v_1^2 + \cdots + v_{n-1}^2)}{2} \right\} \mathrm{d}v_1 \cdots \mathrm{d}v_{n-1},$$

and we can prove by induction the following hypothesis

$$H_n'' : \mathrm{I}_1 = \Phi_{n-1}\left(\left\{ x_{1,n} s_{n,1}^{-1/2} + s_{n,1}^{1/2}/2, \ldots, x_{n-1,n} s_{n,n-1}^{-1/2} + s_{n,n-1}^{1/2}/2 \right\}^\top ; R_1 \right), \quad \forall n \geq 2,$$

where R_1 is the correlation matrix coming from the covariance matrix Σ_1 and is defined by

$$R_1 = \begin{bmatrix} 1 & (s_{n,2}/s_{n,1})^{1/2} & \cdots & (s_{n,n-1}/s_{n,1})^{1/2} \\ (s_{n,2}/s_{n,1})^{1/2} & 1 & \cdots & (s_{n,n-1}/s_{n,2})^{1/2} \\ \vdots & \vdots & \ddots & \vdots \\ (s_{n,n-1}/s_{n,1})^{1/2} & (s_{n,n-1}/s_{n,2})^{1/2} & \cdots & 1 \end{bmatrix}.$$

For $n = 2$, then H_2'':

$$\mathrm{I}_1 = \int_{-\infty}^{x_{1,2} s_{2,1}^{-1/2} + s_{2,1}^{1/2}/2} \phi(v_1) \mathrm{d}v_1 = \Phi\left(x_{1,2} s_{2,1}^{-1/2} + s_{2,1}^{1/2}/2 \right).$$

For $n = 3$, then H_3'':

$$\begin{aligned}
\mathrm{I}_1 &= \int_{-\infty}^{x_{2,3}s_{3,2}^{-1/2}+s_{3,2}^{1/2}/2} \int_{-\infty}^{(x_{1,3}+s_{3,1}/2-v_2s_{3,2})t_{2,1}^{1/2}} 2\pi^{-1} \exp\left\{-\frac{(v_1^2+v_2^2)}{2}\right\} \mathrm{d}v_1\mathrm{d}v_2 \\
&= \int_{-\infty}^{x_{2,3}s_{3,2}^{-1/2}+s_{3,2}^{1/2}/2} (2\pi)^{-1/2} \exp\left\{-\frac{v_2^2}{2}\right\} \Phi\left((x_{1,3}+s_{3,1}/2-v_2s_{3,2})s_{2,1}^{-1/2}\right) \mathrm{d}v_2 \\
&= \Phi_2\left(\left\{x_{2,3}s_{3,2}^{-1/2}+s_{3,2}^{1/2}/2, x_{1,3}s_{3,1}^{-1/2}+s_{3,1}^{1/2}/2\right\}^\top; (s_{3,2}/s_{3,1})^{1/2}\right).
\end{aligned}$$

Assume our hypothesis true for all $2 \leq k \leq n$. Note that our hypothesis H_k'' can re-considered as

$$\begin{aligned}
\mathrm{I}_1 = &\int_{-\infty}^{x_{k-1,k}s_{k,k-1}^{-1/2}+s_{k,k-1}^{1/2}/2} \phi(v_{k-1}) \\
&\times \Phi_{k-2}\left(\left\{\frac{x_{1,k}+s_{k,1}/2-v_{k-1}s_{k,k-1}^{1/2}}{s_{k-1,1}^{1/2}}, \cdots, \frac{x_{k-2,k}+s_{k,k-2}/2-v_{k-1}s_{k,k-1}^{1/2}}{s_{k-1,k-2}^{1/2}}\right\}^\top; R_{k-1}\right) \mathrm{d}v_{k-1}
\end{aligned}$$

where the components of the correlation matrix R_{k-1} are $(s_{k-1,j}/s_{k-1,i})^{1/2}, i, j = 1, \ldots, k-2$.

For H_{k+1}'' we have,

$$\begin{aligned}
\mathrm{I}_1 = &\int_{-\infty}^{x_{k,k+1}s_{k+1,k}^{-1/2}+s_{k+1,k}^{1/2}/2} \int_{-\infty}^{(x_{k-1,k+1}+s_{k+1,k-1}/2-v_ks_{k+1,k})s_{k,k-1}^{-1/2}} \cdots \\
&\int_{-\infty}^{(x_{1,k+1}+s_{k+1,1}/2-\sum_{l=1}^{k-1} v_{k+1-l}s_{k-l+2,k+1-l}^{1/2})s_{2,1}^{-1/2}} (2\pi)^{-\frac{k}{2}} \exp\left\{-\frac{(v_1^2+\cdots+v_k^2)}{2}\right\} \mathrm{d}v_1 \cdots \mathrm{d}v_k.
\end{aligned}$$

For the integrations w.r.t. v_1 up to v_{k-2}, denoting the constants by $c_i = x_{i,k+1} + s_{k+1,i}/2 - v_k s_{k+1,k}^{1/2}, i = 1, \ldots, k-2$ and using the induction hypothesis leads to

$$\begin{aligned}
\mathrm{I}_1 = &\int_{-\infty}^{x_{k,k+1}s_{k+1,k}^{-1/2}+s_{k+1,k}^{1/2}/2} \int_{-\infty}^{(x_{k-1,k+1}+s_{k+1,k-1}/2-v_ks_{k+1,k})s_{k,k-1}^{-1/2}} (2\pi)^{-1} \exp\left\{-\frac{1}{2}(v_{k-1}^2+v_k^2)\right\} \\
&\times \Phi_{k-2}\left(\left\{\left(c_1 - v_k s_{k,k-1}^{1/2}\right) s_{k-1,1}^{-1/2}, \ldots, \left(c_{k-2} - v_k s_{k,k-1}^{1/2}\right) s_{k-1,k-2}^{-1/2}\right\}^\top; R_{k-1}\right) \mathrm{d}v_{k-1}\mathrm{d}v_k.
\end{aligned}$$

Applying Lemma 3.3 a first time yields to

$$\begin{aligned}
\mathrm{I}_1 = &\int_{-\infty}^{x_{k,k+1}s_{k+1,k}^{-1/2}+s_{k+1,k}^{1/2}/2} \phi(v_k) \\
&\times \Phi_{k-2}\left(\left\{c_1 s_{k,1}^{-1/2}, \ldots, c_{k-2}s_{k,k-2}^{-1/2}, \left(x_{k-1,k+1}+s_{k+1,k-1}/2-v_ks_{k+1,k}^{1/2}\right) s_{k,k-1}^{-1/2}\right\}^\top; R_k\right) \mathrm{d}v_k,
\end{aligned}$$

where the components of the correlation matrix R_k are $(s_{k,j}/s_{k,i})^{1/2}, i, j = 1, \ldots, k-1$. Next, a second application of Lemma 3.3 completes the induction proof as it reduces to

$$\mathrm{I}_1 = \Phi_k\left(\left\{x_{1,k+1}s_{k+1,1}^{-1/2}+s_{k+1,1}^{1/2}/2, \ldots, x_{k,k+1}s_{k+1,k}^{-1/2}+s_{k+1,k}^{1/2}/2\right\}^\top; R_{k+1}\right)$$

where the components of the correlation matrix R_{k+1} are $(s_{k+1,j}/s_{k+1,i})^{1/2}, i,j = 1,\dots,k$. Consequently our hypothesis is true for all $n \geq 2$ and hence

$$\mathrm{I}_1 = \Phi_{n-1}\left(\left\{x_{1,n}s_{n,1}^{-1/2} + s_{n,1}^{1/2}/2, \dots, x_{n-1,n}s_{n,n-1}^{-1/2} + s_{n,n-1}^{1/2}/2\right\}^\top; R_n\right), \tag{C.10}$$

where R_n corresponds with the definition of R_1 given earlier. For the second part of (C.6) it is easy to see that

$$\mathrm{I}_2 = \Phi_{p-n}\left(\left\{x_{n+1,n}s_{n+1,n}^{-1/2} + s_{n+1,n}^{1/2}/2, \cdots, x_{p,n}s_{p,n}^{-1/2} + s_{p,n}^{1/2}/2\right\}^\top; R_2\right),$$

where R_2 is the correlation matrix, coming from Σ_2, with coefficients $(s_{n+j,n}/s_{n+i,n})^{1/2}, i,j = 1,\dots,p-n$. Furthermore, let $x_{-i} = x\backslash\{x_i\}$ and $s_{-i} = s\backslash\{s_i\}$ we can conclude that there equivalence between (C.6) and

$$\mathbb{E}[Z] = e^{-x_n}\Phi_{p-1}\left(\left\{(x_{-n} - x_n)|s_{-n} - s_n|^{-1/2} + |s_{-n} - s_n|^{1/2}/2\right\}^\top; R\right)$$

where

$$R = \left[\begin{array}{c:c} R_1 & 0 \\ \hdashline 0 & R_2 \end{array}\right],$$

is a $(p-1)\times(p-1)$ correlation matrix. Clearly it is also equivalent to

$$\mathbb{E}[Z] = e^{-x_n}\Pr\big(\,\epsilon_{n,1} \leq x_{1,n} + \tfrac{1}{2}s_{n,1}, \cdots, \epsilon_{n,n-1} \leq x_{n-1,n} + \tfrac{1}{2}s_{n,n-1}, \\ \epsilon_{n+1,n} \leq x_{n+1,n} + \tfrac{1}{2}s_{n+1,n}, \cdots, \epsilon_{p,n} \leq x_{p,n} + \tfrac{1}{2}s_{p,n}\Big),$$

which completes the proof.

C.3 Proof of Theorem 3.7

Identically as in Appendix C.1, we use the notations $\epsilon_{i,j} = \epsilon(\beta s_i) - \epsilon(\beta s_j)$ and $x_{i,j} = x_i - x_j\ \forall i,j$. Moreover, for conveniency we will suppose here that the s_i's are ordered increasingly. Applying Lemma 2.7 gives

$$\begin{aligned}
&-\log\Pr(\eta'(s_1) \leq e^{x_1}, \dots, \eta'(s_n) \leq e^{x_n}) \\
&\quad= \mathbb{E}\left[\exp\left\{\max\left(\epsilon(\beta s_1) - \tfrac{\beta}{2}s_1 - x_1, \dots, \epsilon(\beta s_n) - \tfrac{\beta}{2}s_n - x_n\right)\right\}\right] \\
&\quad= \mathbb{E}\left[\exp\left\{\epsilon(\beta s_1) - \tfrac{\beta}{2}s_1\right\}\exp\left\{\max\left(-x_1, \epsilon_{2,1} - \tfrac{1}{2}a_{2,1}^2 - x_2, \dots, \epsilon_{n,1} - \tfrac{1}{2}a_{n,1}^2 - x_n\right)\right\}\right] \\
&\quad= \mathbb{E}\left[\exp\left\{\max\left(-x_1, \epsilon_{2,1} - \tfrac{1}{2}a_{2,1}^2 - x_2, \dots, \epsilon_{n,1} - \tfrac{1}{2}a_{n,1}^2 - x_n\right)\right\}\right],
\end{aligned}$$

where $a_{i,j} = (\beta|s_j - s_i|)^{1/2}$. Let us define

$$Y = \max\left(\exp\{-x_1\}, \exp\left\{\epsilon_{2,1} - \tfrac{1}{2}a_{2,1}^2 - x_2\right\}, \ldots, \exp\left\{\epsilon_{n,1} - \tfrac{1}{2}a_{n,1}^2 - x_n\right\}\right).$$

Then it is easily shown that Y takes the value $\exp\{-x_1\}$ with probability

$$\Pr\left(\epsilon_{2,1} \leq x_{2,1} + \tfrac{1}{2}a_{2,1}^2, \cdots, \epsilon_{n,1} \leq x_{n,1} + \tfrac{1}{2}a_{n,1}^2\right)$$

and, for $j = 2, \ldots, n$, the value, $\exp\{\epsilon_{j,1} - \frac{1}{2}a_{j,1}^2 - x_j\}$ with probability

$$\Pr\Big(\epsilon_{j,1} \geq x_{j,1} + \tfrac{1}{2}a_{j,1}^2, \ldots, \epsilon_{j,j-1} \geq x_{j,j-1} + \tfrac{1}{2}a_{j,j-1}^2, \\ \epsilon_{j+1,j} \leq x_{j+1,j} + \tfrac{1}{2}a_{j+1,j}^2, \ldots, \epsilon_{n,j} \leq x_{n,j} + \tfrac{1}{2}a_{n,j}^2\Big).$$

Hence the calculation of the expected value of Y can be considered as the sum of n expectations. The first one is straightforward, if we consider $Y_1 = \exp\{-x_1\}$ that occurs with probability $\Pr\left(\epsilon_{2,1} \leq x_{2,1} + \frac{1}{2}a_{2,1}^2, \ldots, \epsilon_{n,1} \leq x_{n,1} + \frac{1}{2}a_{n,1}^2\right)$ then

$$\mathbb{E}[Y_1] = e^{-x_1}\Phi_{n-1}\left(\left\{x_{2,1}a_{2,1}^{-1} + a_{2,1}/2, \ldots, x_{n,1}a_{n,1}^{-1} + a_{n,1}/2\right\}^\top; R^{(1)}\right)$$

where

$$R^{(1)} = \begin{bmatrix} 1 & a_{2,1}/a_{3,1} & \cdots & a_{2,1}/a_{n,1} \\ a_{2,1}/a_{3,1} & 1 & \cdots & a_{3,1}/a_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{2,1}/a_{n,1} & a_{3,1}/a_{n,1} & \cdots & 1 \end{bmatrix}.$$

For the expected value of the $n-1$ others terms, we only need to apply Lemma 3.6 to obtain the result. Note that $\epsilon_{j,j} = 0, \forall j$ so when $j = n$ the probability of occurrence of Y_n is

$$\Pr\left(\epsilon_{n,1} \geq x_{n,1} + \tfrac{1}{2}a_{n,1}^2, \ldots, \epsilon_{n,n-1} \geq x_{n,n-1} + \tfrac{1}{2}a_{n,n-1}^2\right).$$

For example, for the cases $j = 2$ and $j = n$ we respectively get

$$\mathbb{E}[Y_2] = e^{-x_2}\Phi_{n-1}\left(\left\{x_{1,2}a_{2,1}^{-1} + a_{2,1}/2, \ldots, x_{n,2}a_{n,2}^{-1} + a_{n,2}/2\right\}^\top; R^{(2)}\right)$$

and

$$\mathbb{E}[Y_n] = e^{-x_n}\Phi_{n-1}\left(\left\{x_{1,n}a_{n,1}^{-1} + a_{n,1}/2, \ldots, x_{n-1,n}a_{n,n-1}^{-1} + a_{n,n-1}/2\right\}^\top; R^{(n)}\right),$$

where

$$R^{(2)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & a_{3,2}/a_{4,2} & \cdots & a_{3,2}/a_{n,2} \\ 0 & a_{3,2}/a_{4,2} & 1 & \cdots & a_{4,2}/a_{n,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{3,2}/a_{n,2} & a_{4,2}/a_{n,2} & \cdots & 1 \end{bmatrix}$$

and

$$R^{(n)} = \begin{bmatrix} 1 & a_{n,2}/a_{n,1} & \cdots & a_{n,n-1}/a_{n,1} \\ a_{n,2}/a_{n,1} & 1 & \cdots & a_{n,n-1}/a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,n-1}/a_{n,1} & a_{n,n-1}/a_{n,2} & \cdots & 1 \end{bmatrix}.$$

C.4 Proof of Lemma 3.8

(i) Suppose ϵ_1 and ϵ_2 are standard normally distributed and respectively independent of ϵ and ϵ', then

$$\begin{aligned} \mathbb{E}e^{\epsilon-u/2}\Phi_2\left(\{a\epsilon+b, c\epsilon'+d\}^\top;\rho\right) &= \mathbb{E}_{\epsilon,\epsilon'}\mathbb{E}\left(e^{\epsilon-u/2}\mathbb{1}_{\{\epsilon_1\le a\epsilon+b,\epsilon_2\le c\epsilon'+d\}}|\epsilon,\epsilon'\right) \\ &= \mathbb{E}\, e^{\epsilon-u/2}\mathbb{1}_{\{\epsilon_1\le a\epsilon+b,\epsilon_2\le c\epsilon'+d\}}. \end{aligned}$$

Noting that the joint density of (ϵ,ϵ') is

$$f_{\epsilon,\epsilon'}(x,y) = (2\pi)^{-1}(v(u-v))^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{(x-y)^2}{u-v}+\frac{y^2}{v}\right)\right\},$$

hence by Fubini's Theorem and Lemma 3.3

$$\begin{aligned} &\mathbb{E}\, e^{\epsilon-u/2}\mathbb{1}_{\{\epsilon_1\le a\epsilon+b,\epsilon_2\le c\epsilon'+d\}} \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\mathbb{E}\left(e^{\epsilon-u/2}\mathbb{1}_{\{\epsilon_1\le a\epsilon+b,\epsilon_2\le c\epsilon'+d\}}|\epsilon_1,\epsilon_2\right) \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\int_{\frac{\epsilon_2-d}{c}}^{\infty}\int_{\frac{\epsilon_1-b}{a}}^{\infty}(2\pi)^{-1}(v(u-v))^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{(x-y-(u-v))^2}{u-v}+\frac{(y-v)^2}{v}\right)\right\}\mathrm{d}x\mathrm{d}y \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\int_{-\infty}^{\frac{d-\epsilon_2}{c}}\int_{-\infty}^{\frac{b-\epsilon_1}{a\sqrt{u-v}}}(2\pi)^{-1}v^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{(x\sqrt{u-v}-y+(u-v))^2}{u-v}+\frac{(y+v)^2}{v}\right)\right\}\mathrm{d}x\mathrm{d}y \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\int_{-\infty}^{\frac{d-\epsilon_2}{c}}(2\pi v)^{-1/2}\exp\left\{-\frac{(y+v)^2}{2v}\right\}\Phi\left(\frac{\frac{b-\epsilon_1}{a}-y+u-v}{\sqrt{u-v}}\right)\mathrm{d}y \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\int_{-\infty}^{\frac{d-\epsilon_2}{c}+\sqrt{v}}\phi(y)\Phi\left(\frac{\frac{b-\epsilon_1}{a}-y\sqrt{v}+u-v}{\sqrt{u-v}}\right)\mathrm{d}y \\ &= \mathbb{E}_{\epsilon_1,\epsilon_2}\Phi_2\left(\left\{\frac{b-\epsilon_1+au}{a\sqrt{u}},\frac{d-\epsilon_2+cv}{c\sqrt{v}}\right\}^\top;\sqrt{\frac{u}{v}}\right). \end{aligned}$$

Let ϵ_3 and ϵ_4 be mutually independent of ϵ_1 and ϵ_2, we can now finish the proof of the lemma by a similar trick as above

$$\begin{aligned}
&\mathbb{E}_{\epsilon_1,\epsilon_2}\Phi_2\left(\left\{\frac{b-\epsilon_1+au}{a\sqrt{u}},\frac{d-\epsilon_2+cv}{c\sqrt{v}}\right\}^\top;\sqrt{\frac{u}{v}}\right)\\
&\quad=\mathbb{E}_{\epsilon_1,\epsilon_2}\mathbb{E}\left(\mathbb{1}_{\{\epsilon_3\le\frac{b-\epsilon_1+au}{a\sqrt{u}},\epsilon_4\le\frac{d-\epsilon_2+cv}{c\sqrt{v}}\}}|\epsilon_1,\epsilon_2\right)\\
&\quad=\mathbb{E}\,\mathbb{1}_{\{\epsilon_3\le\frac{b-\epsilon_1+au}{a\sqrt{u}},\epsilon_4\le\frac{d-\epsilon_2+cv}{c\sqrt{v}}\}}\\
&\quad=\Pr\left(\epsilon_3\le\frac{b-\epsilon_1+au}{a\sqrt{u}},\epsilon_4\le\frac{d-\epsilon_2+cv}{c\sqrt{v}}\right)\\
&\quad=\Pr\left(a\sqrt{u}\epsilon_3+\epsilon_1\le b+au,c\sqrt{v}\epsilon_4+\epsilon_2\le d+cv\right)
\end{aligned}$$

Moreover $a\sqrt{u}\epsilon_3+\epsilon_1\sim\mathcal{N}(0,1+a^2u)$ and $c\sqrt{v}\epsilon_4+\epsilon_2\sim\mathcal{N}(0,1+c^2v)$ and their covariance is

$$\operatorname{Cov}(a\sqrt{u}\epsilon_3+\epsilon_1,c\sqrt{v}\epsilon_4+\epsilon_2)=ac\sqrt{u}\sqrt{v}\operatorname{Cov}(\epsilon_3,\epsilon_4)+\operatorname{Cov}(\epsilon_1,\epsilon_2)=acv+\rho,$$

and correlation

$$\rho^*\equiv\operatorname{Corr}(a\sqrt{u}\epsilon_3+\epsilon_1,c\sqrt{v}\epsilon_4+\epsilon_2)=(acv+\rho)((1+a^2u)(1+c^2v))^{-1/2}.$$

Hence the proof is completed as

$$\Pr(a\sqrt{u}\epsilon_3+\epsilon_1\le b+au,c\sqrt{v}\epsilon_4+\epsilon_2\le d+cv)=\Phi_2\left(\left\{\frac{au+b}{\sqrt{1+a^2u}},\frac{cv+d}{\sqrt{1+c^2d}}\right\}^\top;\rho^*\right).$$

(ii) The scheme of the proof is identical to (i) when considering the joint density of (ϵ,ϵ') to be

$$f_{\epsilon,\epsilon'}(x,y)=(2\pi)^{-1}(uv)^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{x^2}{u}+\frac{y^2}{v}\right)\right\}.$$

C.5 Proof of Lemma 3.9

Note that because ϵ is independent of both ϵ' and ϵ''

$$\begin{aligned}
&\mathbb{E}e^{\epsilon-u/2}\Phi_2\left(\left\{a\epsilon+b\epsilon'+c,d\epsilon+e\epsilon''+f\right\}^\top;\rho\right)\\
&\quad=\mathbb{E}_\epsilon e^{\epsilon-u/2}\mathbb{E}_{\epsilon',\epsilon''}\Phi_2\left(\left\{a\epsilon+b\epsilon'+c,d\epsilon+e\epsilon''+f\right\}^\top;\rho\right).
\end{aligned}$$

Furthermore, considering ϵ_1 and ϵ_2 two standard normally distributed random variables such that ϵ_1 is independent of ϵ and ϵ' while ϵ_2 is independent of ϵ and ϵ'', then

$$\begin{aligned}
\mathbb{E}_{\epsilon',\epsilon''}\Phi_2\left(\left\{a\epsilon+b\epsilon'+c,d\epsilon+e\epsilon''+f\right\}^\top;\rho\right)&=\Pr\left(\epsilon_1\le a\epsilon+b\epsilon'+c,\epsilon_2\le d\epsilon+e\epsilon''+f\right)\\
&=\Pr\left(\epsilon_1-b\epsilon'\le a\epsilon+c,\epsilon_2-e\epsilon'\le d\epsilon'+f\right),
\end{aligned}$$

where $\epsilon_1 - b\epsilon' \sim \mathcal{N}(0, 1+b^2v)$, $\epsilon_2 - e\epsilon'' \sim \mathcal{N}(0, 1+e^2w)$ and

$$\operatorname{Cov}(\epsilon_1 - b\epsilon', \epsilon_2 - e\epsilon'') = \operatorname{Cov}(\epsilon_1, \epsilon_2) + be \operatorname{Cov}(\epsilon', \epsilon' + \epsilon''') = \rho + bev.$$

Hence

$$\Pr(\epsilon_1 - b\epsilon' \leq a\epsilon + c, \epsilon_2 - e\epsilon' \leq d\epsilon' + f) = \Phi_2\left(\left\{\frac{a\epsilon + c}{\sqrt{1+b^2v}}, \frac{d\epsilon + f}{\sqrt{1+e^2w}}\right\}^\top; \rho'\right)$$

where $\rho' = \frac{\rho + bev}{\sqrt{1+b^2v}\sqrt{1+e^2w}}$ and using Lemma 3.3, it leads to

$$\begin{aligned}
&\mathbb{E}_\epsilon e^{\epsilon - u/2} \mathbb{E}_{\epsilon', \epsilon''} \Phi_2\left(\{a\epsilon + b\epsilon' + c, d\epsilon + e\epsilon'' + f\}^\top; \rho\right) \\
&\qquad = \Phi_2\left(\left\{\frac{a\epsilon + c}{\sqrt{1+b^2v}}, \frac{d\epsilon + f}{\sqrt{1+e^2w}}\right\}^\top; \rho'\right) \\
&\qquad = \int_{-\infty}^{\infty} e^{x - \frac{u}{2}} \Phi_2\left(\left\{\frac{ax + c}{\sqrt{1+b^2v}}, \frac{dx + f}{\sqrt{1+e^2w}}\right\}^\top; \rho'\right) (2\pi u)^{-1/2} \exp\left\{-\frac{x^2}{2u}\right\} \mathrm{d}x \\
&\qquad = \int_{-\infty}^{\infty} (2\pi u)^{-1/2} \exp\left\{-\frac{(x-u)^2}{2u}\right\} \Phi_2\left(\left\{\frac{ax + c}{\sqrt{1+b^2v}}, \frac{dx + f}{\sqrt{1+e^2w}}\right\}^\top; \rho'\right) \mathrm{d}x \\
&\qquad = \int_{-\infty}^{\infty} \phi(x) \Phi_2\left(\left\{\frac{ax\sqrt{u} + au + c}{\sqrt{1+b^2v}}, \frac{dx\sqrt{u} + du + f}{\sqrt{1+e^2w}}\right\}^\top; \rho'\right) \mathrm{d}x \\
&\qquad = \Phi_2\left(\left\{\frac{au + c}{\sqrt{1+a^2u+b^2v}}, \frac{du + f}{\sqrt{1+d^2u+e^2w}}\right\}^\top; \rho^*\right),
\end{aligned}$$

where $\rho^* = (\rho + adu + bev)((1+a^2u+b^2v)(1+d^2u+e^2w))^{-1/2}$.

C.6 Proof of Lemma 3.10

The property of mutual independence between ϵ, ϵ' and ϵ'' allows to write

$$\begin{aligned}
&\mathbb{E}\, e^{\epsilon + \epsilon' - (u+v)/2} \Phi_2\Big(\{a\epsilon + b\epsilon' + c\epsilon'' + d, e\epsilon + f\}^\top; \rho\Big) \\
&\qquad = \mathbb{E}_\epsilon e^{\epsilon - u/2} \mathbb{E}_{\epsilon'} e^{\epsilon' - v/2} \mathbb{E}_{\epsilon''} \Phi_2\left(\left(\{a\epsilon + b\epsilon' + c\epsilon'' + d, e\epsilon + f\}^\top; \rho\right)\right).
\end{aligned}$$

When introducing ϵ_1 and ϵ_2 two standard normal random variable, ϵ_1 independent of ϵ, ϵ' and ϵ'' while ϵ_2 is independent of ϵ, the expectation with respect to ϵ'' is given by

$$\begin{aligned}
&\mathbb{E}_{\epsilon''}\, e^{\epsilon + \epsilon' - (u+v)/2} \Phi_2\left(\{a\epsilon + b\epsilon' + c\epsilon'' + d, e\epsilon + f\}^\top; \rho\right) \\
&\qquad = \Pr(\epsilon_1 \leq a\epsilon + b\epsilon' + c\epsilon'' + d, \epsilon_2 \leq e\epsilon + f) \\
&\qquad = \Pr(\epsilon_1 - c\epsilon'' \leq a\epsilon + b\epsilon' + d, \epsilon_2 \leq e\epsilon + f) \\
&\qquad = \Phi_2\left(\left\{\frac{a\epsilon + b\epsilon' + d}{\sqrt{1+c^2w}}, e\epsilon + f\right\}^\top; \frac{\rho}{\sqrt{1+c^2w}}\right).
\end{aligned}$$

as $\epsilon_1 - c\epsilon'' \sim \mathcal{N}(0, 1 + c^2 w)$, $\epsilon_2 \sim \mathcal{N}(0,1)$ and $\mathrm{Cov}(\epsilon_1 - c\epsilon'', \epsilon_2) = \mathrm{Cov}(\epsilon_1, \epsilon_2) = \rho$.

Now a second step consist in calculating the expected value with respect to ϵ', using Lemma 3.3 it leads to,

$$\begin{aligned}
&\mathbb{E}_{\epsilon'} e^{\epsilon' - v/2} \Phi_2 \left(\left\{ \frac{a\epsilon + b\epsilon' + d}{\sqrt{1 + c^2 w}}, e\epsilon + f \right\}^\top ; \frac{\rho}{\sqrt{1 + c^2 w}} \right) \\
&\quad = \int_{-\infty}^{\infty} e^{x - v/2} \Phi_2 \left(\left\{ \frac{a\epsilon + bx + d}{\sqrt{1 + c^2 w}}, e\epsilon + f \right\}^\top ; \frac{\rho}{\sqrt{1 + c^2 w}} \right) (2\pi v)^{-1/2} \exp\left\{ -\frac{x^2}{2v} \right\} \mathrm{d}x \\
&\quad = \int_{-\infty}^{\infty} \phi(x) \Phi_2 \left(\left\{ \frac{a\epsilon + bx\sqrt{v} + bv + d}{\sqrt{1 + c^2 w}}, e\epsilon + f \right\}^\top ; \frac{\rho}{\sqrt{1 + c^2 w}} \right) \mathrm{d}x \\
&\quad = \Phi_2 \left(\left\{ \frac{au + bv + d}{\sqrt{1 + a^2 u + b^2 v + c^2 w}}, \frac{eu + f}{\sqrt{1 + e^2 u}} \right\}^\top ; \frac{\rho}{\sqrt{1 + b^2 v + c^2 w}} \right).
\end{aligned}$$

And finally for the expectation with respect to ϵ we repeat the same method and use Lemma 3.3 again, we get

$$\begin{aligned}
&\mathbb{E}_{\epsilon} e^{\epsilon - u/2} \Phi_2 \left(\left\{ \frac{a\epsilon + bv + d}{\sqrt{1 + b^2 v + c^2 w}}, e\epsilon + f \right\}^\top ; \frac{\rho}{\sqrt{1 + b^2 v + c^2 w}} \right) \\
&= \int_{-\infty}^{\infty} (2\pi u)^{-1/2} \exp\left\{ -\frac{(y-u)^2}{2u} \right\} \Phi_2 \left(\left\{ \frac{ay + bv + d}{\sqrt{1 + b^2 v + c^2 w}}, ey + f \right\}^\top ; \frac{\rho}{\sqrt{1 + b^2 v + c^2 w}} \right) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \phi(y) \Phi_2 \left(\left\{ \frac{ay\sqrt{u} + au + bv + d}{\sqrt{1 + b^2 v + c^2 w}}, ey\sqrt{u} + eu + f \right\}^\top ; \frac{\rho}{\sqrt{1 + b^2 v + c^2 w}} \right) \mathrm{d}y \\
&= \Phi_2 \left(\left\{ \frac{au + bv + d}{\sqrt{1 + a^2 u + b^2 v + c^2 w}}, \frac{eu + f}{\sqrt{1 + e^2 u}} \right\}^\top ; \frac{\rho + aeu}{\sqrt{1 + a^2 u + b^2 v + c^2 w}\sqrt{1 + e^2 u}} \right).
\end{aligned}$$

C.7 Proof of Theorem 3.11

The scheme of the proof is similar to de Haan and Ferreira (2006, Proposition 2.1). Results for the univariate processes presented in Theorem 3.5 will also be used. For clarity purposes $\beta s_i^{(j)}$ will simply be denoted by $s_i^{(j)}$, for $i, = 1, 2, 3; j = 1, 2$. Again, based on Lemma 2.7 we can give the following decomposition:

$$\begin{aligned}
&-\log \Pr(\eta(s_1) \le e^{x_1}, \eta(s_2) \le e^{x_2}, \eta(s_3) \le e^{x_3}) \\
&= \mathbb{E}\left[\max\left(e^{\epsilon_1\left(s_1^{(1)}\right) + \epsilon_2\left(s_1^{(2)}\right) - \frac{\left|s_1^{(1)}\right| + \left|s_1^{(2)}\right|}{2} - x_1}, e^{\epsilon_1\left(s_2^{(1)}\right) + \epsilon_2\left(s_2^{(2)}\right) - \frac{\left|s_2^{(1)}\right| + \left|s_2^{(2)}\right|}{2} - x_2}, \right.\right. \\
&\qquad \left.\left. e^{\epsilon_1\left(s_3^{(1)}\right) + \epsilon_2\left(s_3^{(2)}\right) - \frac{\left|s_3^{(1)}\right| + \left|s_3^{(2)}\right|}{2} - x_3} \right) \right] \\
&= P_1 + P_2 + P_3,
\end{aligned}$$

where this three components are

$$P_1 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_1^{(1)}\right)-\left|s_1^{(1)}\right|/2-x_1} \mathbb{E}_{\epsilon_2}\left[\max\left(e^{\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)+\epsilon_2\left(s_3^{(2)}\right)-\frac{\left|s_3^{(1)}\right|-\left|s_1^{(1)}\right|+\left|s_3^{(2)}\right|}{2}-x_3+x_1}, \right.\right.\right.$$
$$\left.\left.\left. e^{\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)+\epsilon_2\left(s_2^{(2)}\right)-\frac{\left|s_2^{(1)}\right|-\left|s_1^{(1)}\right|+\left|s_2^{(2)}\right|}{2}-x_2+x_1}, e^{\epsilon_2\left(s_1^{(2)}\right)-\left|s_1^{(2)}\right|/2} \right)\right] \middle| \epsilon_1 \right]$$
$$= \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_1^{(1)}\right)-\left|s_1^{(1)}\right|/2-x_1} \Phi_2\left(\left\{ \frac{\sqrt{h_{1,2}^{(2)}}}{2} + \frac{x_2 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{\left|s_1^{(1)}\right|-\left|s_2^{(1)}\right|}{2}}{\sqrt{h_{1,2}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left. \frac{\sqrt{h_{1,3}^{(2)}}}{2} + \frac{x_3 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{\left|s_1^{(1)}\right|-\left|s_3^{(1)}\right|}{2}}{\sqrt{h_{1,3}^{(2)}}} \right\}^\top ; \rho_{2,3}^{(1)} \right) \right], \tag{C.11}$$

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-\left|s_2^{(1)}\right|/2-x_2} \mathbb{E}_{\epsilon_2}\left[\max\left(e^{\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\epsilon_2\left(s_3^{(2)}\right)-\frac{\left|s_3^{(1)}\right|-\left|s_2^{(1)}\right|+\left|s_3^{(2)}\right|}{2}-x_3+x_2}, \right.\right.\right.$$
$$\left.\left.\left. e^{\epsilon_2\left(s_2^{(2)}\right)-\left|s_2^{(2)}\right|/2}, e^{\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\epsilon_2\left(s_1^{(2)}\right)-\frac{\left|s_1^{(1)}\right|-\left|s_2^{(1)}\right|+\left|s_1^{(2)}\right|}{2}-x_1+x_2} \right)\right] \middle| \epsilon_1 \right]$$
$$= \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-\left|s_2^{(1)}\right|/2-x_2} \Phi_2\left(\left\{ \frac{\sqrt{h_{1,2}^{(2)}}}{2} + \frac{x_1 - x_2 + \epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right) - \frac{\left|s_2^{(1)}\right|-\left|s_1^{(1)}\right|}{2}}{\sqrt{h_{1,2}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left. \frac{\sqrt{h_{2,3}^{(2)}}}{2} + \frac{x_3 - x_2 + \epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{\left|s_2^{(1)}\right|-\left|s_3^{(1)}\right|}{2}}{\sqrt{h_{2,3}^{(2)}}} \right\}^\top ; \rho_{1,3}^{(2)} \right) \right], \tag{C.12}$$

and

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-\left|s_3^{(1)}\right|/2-x_3}\mathbb{E}_{\epsilon_2}\left[\max\left(e^{\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\epsilon_2\left(s_2^{(2)}\right)-\frac{\left|s_2^{(1)}\right|-\left|s_3^{(1)}\right|+\left|s_2^{(2)}\right|}{2}-x_2+x_3},\right.\right.\right.$$
$$\left.\left.\left.e^{\epsilon_2\left(s_3^{(2)}\right)-\left|s_3^{(2)}\right|/2}, e^{\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\epsilon_2\left(s_1^{(2)}\right)-\frac{\left|s_1^{(1)}\right|-\left|s_3^{(1)}\right|+\left|s_1^{(2)}\right|}{2}-x_1+x_3}\right)\right]\middle|\epsilon_1\right]$$
$$= \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-\left|s_3^{(1)}\right|/2-x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_1-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{\left|s_3^{(1)}\right|-\left|s_1^{(1)}\right|}{2}}{\sqrt{h_{1,3}^{(2)}}},\right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_2-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_2\left(s_2^{(1)}\right)-\frac{\left|s_3^{(1)}\right|-\left|s_2^{(1)}\right|}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^{\top};\rho_{1,2}^{(3)}\right)\right]. \tag{C.13}$$

Note that the correlation coefficients $\rho_{j,k}^{i}$, $i,j,k = 1,2,3$, are defined in Theorem 3.5 when replacing $a_{i,j}$ by $\sqrt{\beta h_{i,j}^{(2)}}$. In order to evaluate P_1, P_2 and P_3 different scenarios, depending on the on the values of $s_1^{(1)}, s_2^{(1)}$ and $s_3^{(1)}$, need to be investigated. Each of them are considered below.

Case 1: $0 \leq s_1^{(1)} \leq s_2^{(1)} \leq s_3^{(1)}$**.**

First of all, using the notation introduced earlier on, we recall that here $u_j^{(1)} - u_i^{(1)}$ can be replaced by $h_{i,j}^{(1)}$. In (C.11), we remark that $e^{\epsilon_1\left(s_1^{(1)}\right)-s_1^{(1)}/2-x_1}$, is independent of the other part and furthermore we know that $\mathbb{E}\left[e^{\epsilon_1\left(s_1^{(1)}\right)-s_1^{(1)}/2}\right] = 1$. Thus (C.11) is equal to

$$P_1 = e^{-x_1}\mathbb{E}_{\epsilon_1}\left[\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_2-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_3-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}\right\}^{\top};\rho_{2,3}^{(1)}\right)\right] \tag{C.14}$$

Suppose N_1 and N_2 are two standard normal random variables independent of $\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$ such that

$$P_1 = e^{-x_1}\Pr\left(\sqrt{h_{1,2}^{(2)}}N_1+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)\le x_2-x_1+\frac{1}{2}\left(h_{1,2}^{(1)}+h_{1,2}^{(2)}\right),\right.$$
$$\left.\sqrt{h_{1,3}^{(2)}}N_2+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)\le x_3-x_1+\frac{1}{2}\left(h_{1,3}^{(1)}+h_{1,3}^{(2)}\right)\right).$$

Moreover, it is easy to show that

$$\begin{aligned} A &= \sqrt{h_{1,2}^{(2)}}N_1+\epsilon_2\left(s_1^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)\sim\mathcal{N}\left(0,h_{1,2}^{(1)}+h_{1,2}^{(2)}\right),\\ B &= \sqrt{h_{1,3}^{(2)}}N_2+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)\sim\mathcal{N}\left(0,h_{1,3}^{(1)}+h_{1,3}^{(2)}\right),\\ \mathrm{Cov}(A,B) &= \sqrt{h_{1,2}^{(2)}}\sqrt{h_{1,3}^{(2)}}\rho_{1,3}^2+h_{1,2}^{(1)}, \end{aligned}$$

which allows us to conclude that

$$P_1 = e^{-x_1}\Phi_2\left(\left\{\frac{g_{1,2}}{2}+\frac{x_2-x_1}{g_{1,2}},\frac{g_{1,2}}{2}+\frac{x_3-x_1}{g_{1,2}}\right\}^\top;\rho_1^*\right), \tag{C.15}$$

where

$$\rho_1^* = \begin{cases} h_{1,2}^{(1)}(g_{1,2}g_{1,3})^{-1} & \text{if } s_2^{(2)}\le s_1^{(2)}\le s_3^{(2)} \text{ or } u_3^{(2)}\le s_1^{(2)}\le s_2^{(2)}\\ \left(\min\left(h_{1,2}^{(1)},h_{1,3}^{(2)}\right)+h_{1,2}^{(1)}\right)(g_{1,2}g_{1,3})^{-1} & \text{otherwise}\end{cases}. \tag{C.16}$$

Note that because $\mathbb{E}\left[e^{\epsilon_1\left(s_1^{(1)}\right)-s_1^{(1)}/2}\right]=1$ and using the fact that $\epsilon_1\left(s_1^{(1)}\right)$, $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)$ are mutually independent, then

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}-x_2}\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_1-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_3-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,3}^{(2)}\right)\right].$$

Then applying Lemma 3.8 (ii), we get

$$P_2 = e^{-x_2}\Phi_2\left(\left\{\frac{g_{1,2}}{2}+\frac{x_1-x_2}{g_{1,2}},\frac{g_{2,3}}{2}+\frac{x_3-x_2}{g_{2,3}}\right\}^\top;\rho_2^*\right), \tag{C.17}$$

where

$$\rho_2^* = \begin{cases} 0 & \text{if } s_2^{(2)} \leq s_1^{(2)} \leq s_3^{(2)} \text{ or } s_3^{(2)} \leq s_1^{(2)} \leq s_2^{(2)} \\ \min\left(h_{1,2}^{(2)}, h_{2,3}^{(2)}\right)(g_{1,2}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.18}$$

Finally, as $\epsilon_1\left(s_1^{(1)}\right)$ is independent of both $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$, then

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}-x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2} + \frac{x_1 - x_3 + \epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right) - \frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2} + \frac{x_2 - x_3 + \epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top ; \rho_{1,2}^{(3)}\right)\right].$$

Then applying Lemma 3.8 (i), we get

$$P_3 = e^{-x_3}\Phi_2\left(\left\{\frac{g_{1,3}}{2} + \frac{x_1 - x_3}{g_{1,3}}, \frac{g_{2,3}}{2} + \frac{x_2 - x_3}{g_{2,3}}\right\}^T ; \rho_3^*\right), \tag{C.19}$$

where

$$\rho_3^* = \begin{cases} h_{2,3}^{(1)}(g_{1,3}g_{2,3})^{-1} & \text{if } s_2^{(2)} \leq s_1^{(2)} \leq s_3^{(2)} \text{ or } s_3^{(2)} \leq s_1^{(2)} \leq s_2^{(2)} \\ \left(\min\left(h_{1,3}^{(2)}, h_{2,3}^{(2)}\right) + h_{2,3}^{(1)}\right)(g_{1,3}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.20}$$

Case 2: $0 \leq s_1^{(1)} \leq s_3^{(1)} \leq s_2^{(1)}$**.**

Observe that here $s_2^{(1)} - s_1^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_3^{(1)} - s_1^{(1)}$ by $h_{1,3}^{(1)}$ and $s_2^{(1)} - s_3^{(1)}$ by $h_{2,3}^{(1)}$. Furthermore $\epsilon_1\left(s_1^{(1)}\right)$ is independent of both $\epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right)$ leading to the equality between P_1 and (C.14). Identical calculations as in Case 1 leads to the conclusion that here P_1 is equal to (C.15) with

$$\rho_1^* = \begin{cases} h_{1,3}^{(1)}(g_{1,2}g_{1,3})^{-1} & \text{if } s_2^{(2)} \leq s_1^{(2)} \leq s_3^{(2)} \text{ or } s_3^{(2)} \leq s_1^{(2)} \leq s_2^{(2)} \\ \left(\min\left(h_{1,2}^{(2)}, h_{1,3}^{(2)}\right) + h_{1,3}^{(1)}\right)(g_{1,2}g_{1,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.21}$$

Note that $\mathbb{E}\left[e^{\epsilon_1\left(s_1^{(1)}\right)-s_1^{(1)}/2}\right] = 1$ and using the fact that $\epsilon_1\left(s_1^{(1)}\right)$ is independent of both $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$, then

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}-x_2}\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_1-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_3-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)-\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,3}^{(2)}\right)\right].$$

Applying Lemma 3.8 (i) leads to (C.17) where

$$\rho_2^* = \begin{cases} h_{2,3}^{(1)}(g_{1,2}g_{2,3})^{-1} & \text{if } s_2^{(2)} \le s_1^{(2)} \le s_3^{(2)} \text{ or } s_3^{(2)} \le s_1^{(2)} \le s_2^{(2)} \\ \left(\min\left(h_{1,2}^{(2)}, h_{2,3}^{(2)}\right)+h_{2,3}^{(1)}\right)(g_{1,2}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.22}$$

Finally as $\epsilon_1\left(s_1^{(1)}\right)$, $\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$ are mutually independent, then

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}-x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_1-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_2-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,2}^{3}\right)\right].$$

Applying Lemma 3.8 (ii) leads to (C.19) with

$$\rho_3^* = \begin{cases} 0 & \text{if } s_2^{(2)} \le s_1^{(2)} \le s_3^{(2)} \text{ or } s_3^{(2)} \le s_1^{(2)} \le s_2^{(2)} \\ \min\left(h_{1,3}^{(2)}, h_{2,3}^{(2)}\right)(g_{1,3}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.23}$$

Case 3: $0 \le s_2^{(1)} \le s_1^{(1)} \le s_3^{(1)}$**.**

Here we observe that $s_1^{(1)} - s_2^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_3^{(1)} - s_1^{(1)}$ by $h_{1,3}^{(1)}$ and $s_3^{(1)} - s_2^{(1)}$ by $h_{2,3}^{(1)}$. Note that $\mathbb{E}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-s_2^{(1)}/2}\right] = 1$ and using the fact that $\epsilon_1\left(s_2^{(1)}\right)$, $\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)$

and $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ are mutually independent, then

$$P_1 = e^{\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}-x_1}\mathbb{E}_{\epsilon_1}\left[\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_2-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_3-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}\right\}^\top;\rho_{2,3}^{(1)}\right)\right].$$

Applying Lemma 3.8 (ii) leads to (C.15) with

$$\rho_1^* = \begin{cases} 0 & \text{if } s_2^{(2)} \le s_1^{(2)} \le s_3^{(2)} \text{ or } s_3^{(2)} \le s_1^{(2)} \le s_2^{(2)} \\ \min\left(h_{1,2}^{(2)}, h_{1,3}^{(2)}\right)(g_{1,2}g_{1,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.24}$$

In the current setup $\epsilon_1\left(s_2^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ leading to

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{-x_2}\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_1-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)+\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_3-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)+\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,3}^{(2)}\right)\right]. \tag{C.25}$$

An identical method to the calculations of P_1 in Case 1 and Case 2 leads to the conclusion that P_2 is here equal to (C.17) where

$$\rho_2^* = \begin{cases} h_{1,2}^{(1)}(g_{1,2}g_{2,3})^{-1} & \text{if } s_2^{(2)} \le s_1^{(2)} \le s_3^{(2)} \text{ or } s_3^{(2)} \le s_1^{(2)} \le s_2^{(2)} \\ \left(\min\left(h_{1,2}^{(2)}, h_{2,3}^{(2)}\right) + h_{1,2}^{(1)}\right)(g_{1,2}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.26}$$

Noting that $\epsilon_1\left(s_2^{(1)}\right)$ is independent of both $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$, it is equal to

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)-\frac{1}{2}h_{2,3}^{(1)}-x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_1-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}},\right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_2-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)-\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,2}^{(3)}\right)\right].$$

Then applying Lemma 3.8 (i) leads to the equality between P_3 and (C.19) where

$$\rho_3^* = \begin{cases} h_{1,3}^{(1)}(g_{1,3}g_{2,3})^{-1} & \text{if } s_2^{(2)} \leq s_1^{(2)} \leq s_3^{(2)} \text{ or } s_3^{(2)} \leq s_1^{(2)} \leq s_2^{(2)} \\ \left(\min\left(h_{1,3}^{(2)}, h_{2,3}^{(2)}\right) + h_{1,3}^{(1)}\right)(g_{1,3}g_{2,3})^{-1} & \text{otherwise} \end{cases}. \tag{C.27}$$

Case 4: $0 \leq s_2^{(1)} \leq s_3^{(1)} \leq s_1^{(1)}$**.**

Observe that here $s_1^{(1)} - s_2^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_1^{(1)} - s_3^{(1)}$ by $h_{1,3}^{(1)}$ and $s_3^{(1)} - s_2^{(1)}$ by $h_{2,3}^{(1)}$. We know that $\epsilon_1\left(s_2^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$, hence

$$P_1 = e^{\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{1}{2}h_{1,2}^{(1)} - x_1}\mathbb{E}_{\epsilon_1}\left[\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2} + \frac{x_2 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{1,3}^{(2)}}}{2} + \frac{x_3 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{1}{2}h_{1,3}^{(1)}}{\sqrt{h_{1,3}^{(2)}}}\right\}^\top ; \rho_{2,3}^{(1)}\right)\right].$$

Applying Lemma 3.8 (i) leads to (C.15) with ρ_1^* defined as in (C.21). Moreover $\epsilon_1\left(s_2^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$, and $\mathbb{E}\left[e^{\epsilon_1\left(s_2^{(1)}\right) - s_2^{(1)}/2 - x_1}\right] = 1$, then P_2 is identical to (C.25). Calculations are then similar to the one in Case 3 and the result obtained here is identical to (C.17) where ρ_2^* is defined as in (C.22). Finally, we also have $\epsilon_1\left(s_2^{(1)}\right)$, $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ and $\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ mutually independent which yields to

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{h_{2,3}^{(1)}}{2} - x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2} + \frac{x_1 - x_3 + \epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right) + \frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2} + \frac{x_2 - x_3 + \epsilon_1\left(s_3^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top ; \rho_{1,2}^{(3)}\right)\right].$$

Then applying Lemma 3.8 (ii) we get that P_3 is equal to (C.19) where ρ_3^* is defined as in (C.23).

Case 5: $0 \leq s_3^{(1)} \leq s_1^{(1)} \leq s_2^{(1)}$**.**

Observe that here $s_2^{(1)} - s_1^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_1^{(1)} - s_3^{(1)}$ by $h_{1,3}^{(1)}$ and $s_2^{(1)} - s_3^{(1)}$ by $h_{2,3}^{(1)}$. Moreover $\epsilon_1\left(s_3^{(1)}\right)$, $\epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ are mutually independent

and thus

$$P_1 = e^{\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}-x_1}\mathbb{E}_{\epsilon_1}\left[\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_2-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\frac{1}{2}h_{1,2}^{(1)}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_3-x_1+\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}\right\}^\top;\rho_{2,3}^{(1)}\right)\right].$$

Applying Lemma 3.8 (ii) leads to (C.15) where ρ_1^* is defined as in (C.24).

We also know that $\epsilon_1\left(s_3^{(1)}\right)$ is independent of both $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$. Thus we have

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)-\frac{h_{1,3}^{(1)}}{2}-x_2}\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2}+\frac{x_1-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)-\frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_3-x_2+\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)-\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,3}^{(2)}\right)\right],$$

to which we can apply Lemma 3.8 (i). It yields to (C.17) where ρ_2^* is defined as in (C.26).

Finally noting that $\epsilon_1\left(s_3^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right)-\epsilon_1\left(s_3^{(1)}\right)$ and $\mathbb{E}\left[e^{\epsilon_1\left(s_3^{(1)}\right)-s_3^{(1)}/2}\right]=1$, it leads that

$$P_3 = \mathbb{E}_{\epsilon_1}\left[e^{-x_3}\Phi_2\left(\left\{\frac{\sqrt{h_{1,3}^{(2)}}}{2}+\frac{x_1-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_1^{(1)}\right)+\frac{1}{2}h_{1,3}^{(1)}}{\sqrt{h_{1,3}^{(2)}}},\right.\right.$$
$$\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2}+\frac{x_2-x_3+\epsilon_1\left(s_3^{(1)}\right)-\epsilon_1\left(s_2^{(1)}\right)+\frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top;\rho_{1,2}^{(3)}\right)\right]. \tag{C.28}$$

The method used here is similar to the one of P_1 in Case 1. We obtain (C.19) where ρ_3^* is defined as in (C.27).

Case 6: $0 \leq s_3^{(1)} \leq s_2^{(1)} \leq s_1^{(1)}$.

Observe that here $s_1^{(1)} - s_2^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_1^{(1)} - s_3^{(1)}$ by $h_{1,3}^{(1)}$ and $s_2^{(1)} - s_3^{(1)}$ by $h_{2,3}^{(1)}$. Moreover $\epsilon_1\left(s_3^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$. Then

$$P_1 = e^{\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{1}{2}h_{1,3}^{(1)} - x_1} \mathbb{E}_{\epsilon_1}\left[\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2} + \frac{x_2 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right) - \frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{1,3}^{(2)}}}{2} + \frac{x_3 - x_1 + \epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{h_{1,3}^{(1)}}{2}}{\sqrt{h_{1,3}^{(2)}}}\right\}^\top ; \rho_{2,3}^{(1)}\right)\right].$$

Applying Lemma 3.8 (i) leads to (C.15) where ρ_1^* is defined as in (C.16). Note that also $\epsilon_1\left(s_3^{(1)}\right)$, $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_2^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ are mutually independent, leading to

$$P_2 = \mathbb{E}_{\epsilon_1}\left[e^{\epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{h_{1,3}^{(1)}}{2} - x_2}\Phi_2\left(\left\{\frac{\sqrt{h_{1,2}^{(2)}}}{2} + \frac{x_1 - x_2 + \epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_1^{(1)}\right) + \frac{h_{1,2}^{(1)}}{2}}{\sqrt{h_{1,2}^{(2)}}}, \right.\right.\right.$$
$$\left.\left.\left.\frac{\sqrt{h_{2,3}^{(2)}}}{2} + \frac{x_3 - x_2 + \epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right) - \frac{h_{2,3}^{(1)}}{2}}{\sqrt{h_{2,3}^{(2)}}}\right\}^\top ; \rho_{1,3}^{(2)}\right)\right].$$

Applying Lemma 3.8 (ii) leads to (C.17) where ρ_2^* is defined as in (C.18).

Finally $\epsilon_1\left(s_3^{(1)}\right)$ is independent of both $\epsilon_1\left(s_1^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ and $\epsilon_1\left(s_2^{(1)}\right) - \epsilon_1\left(s_3^{(1)}\right)$ and also $\mathbb{E}\left[e^{\epsilon_1\left(s_3^{(1)}\right) - s_3^{(1)}/2}\right] = 1$. This yields to the quality between P_3 and (C.28). The method is then similar than in Case 5 and produces (C.19) with ρ_3^* defined as in (C.20).

Case 7: $s_1^{(1)} \leq 0 \leq s_2^{(1)} \leq s_3^{(1)}$.

Applying Lemma 3.9 to (C.11) and (C.12) and Lemma 3.10 to (C.13), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 1.

Case 8: $s_1^{(1)} \leq 0 \leq s_3^{(1)} \leq s_2^{(1)}$.

Applying Lemma 3.9 to (C.11) and (C.13) and Lemma 3.10 to (C.12), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 2.

Case 9: $s_2^{(1)} \leq 0 \leq s_1^{(1)} \leq s_3^{(1)}$**.**

Here we observe that $s_1^{(1)} - s_2^{(1)}$ can be replaced by $h_{1,2}^{(1)}$, $s_3^{(1)} - s_1^{(1)}$ by $h_{1,3}^{(1)}$ and $s_3^{(1)} - s_2^{(1)}$ by $h_{2,3}^{(1)}$. Applying Lemma 3.9 to (C.11) and (C.12) and Lemma 3.10 to (C.13), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 3.

Case 10: $s_2^{(1)} \leq 0 \leq s_3^{(1)} \leq s_1^{(1)}$**.**

Applying Lemma 3.9 to (C.12) and (C.13) and Lemma 3.10 to (C.11), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 2.

Case 11: $s_3^{(1)} \leq 0 \leq s_1^{(1)} \leq s_2^{(1)}$**.**

Applying Lemma 3.9 to (C.11) and (C.13) and Lemma 3.10 to (C.12), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 3.

Case 12: $s_3^{(1)} \leq 0 \leq s_2^{(1)} \leq s_1^{(1)}$**.**

Applying Lemma 3.9 to (C.12) and (C.13) and Lemma 3.10 to (C.11), proves that P_1, P_2 and P_3 are respectively equal to (C.15), (C.17) and (C.19) where ρ_1^*, ρ_2^* and ρ_3^* are defined as in Case 1.

There are $4! = 24$ cases to be treated, however the remaining twelve cases can be calculated by symmetry. Actually, looking at the first twelve cases detailed here, it can be seen that there are only three different possible expressions for each correlation coefficient. Merging them into a single expression gives Theorem 3.11.

Appendix D

Chapter 4: proofs and computational details

D.1 Proofs

D.1.1 Proof of Proposition 4.2

Items (1)–(3) are easily derived following the proof of Propositions (1)–(4) of Arellano-Valle and Genton (2010) and taking into account the next result.

Lemma D.1. *Let* $Y = (Y_1^\top, Y_2^\top)^\top \sim \mathcal{T}_d(\mu, \Omega, \kappa, \nu)$, *where* $Y_1 \in \mathbb{R}$ *and* $Y_2 \in \mathbb{R}^{d-1}$ *with the corresponding partition of the parameters* (μ, Ω, ν) *and* $\kappa = (\kappa_1, 0^\top)^\top$ *with* $\kappa_1 \in \mathbb{R}$. *Then,*

$$(Y_1|Y_2 = y_2) \sim \mathcal{T}(\mu_{1\cdot 2}, \Omega_{11\cdot 2}, \kappa_{1\cdot 2}, \nu_{1\cdot 2}), \qquad y_2 \in \mathbb{R}^{d-1}$$

where $\mu_{1\cdot 2} = \mu_1 + \Omega_{12}\Omega_{22}^{-1}(y_2 - \mu_2)$, $\Omega_{1\cdot 2} = \zeta_2\Omega_{11\cdot 2}$, $\zeta_2 = \{\nu + Q_{\Omega_{22}^{-1}}(z_2)\}/(\nu + d_2)$, $z_2 = \omega_2^{-1}(y_2 - \mu_2)/\Omega_2$, $\omega_2 = diag(\Omega_{22})^{1/2}$, $\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, $\kappa_{1\cdot 2} = \zeta_2^{-1/2}\kappa$, $\nu_{1\cdot 2} = \nu + d - 1$

Proof of Lemma D.1. The marginal density of Y_2 is equal to

$$f_{Y_2}(y_2) = \int_0^\infty \frac{v^{\nu/2-1}e^{-v}}{\Gamma(\nu/2)}\phi_{d-1}\left(\frac{y_2 - \mu_2}{\sqrt{\frac{\nu}{2v}}}; \Omega_{22}\right)\left(\frac{2v}{\nu}\right)^{(d-1)/2} \mathrm{d}v = \psi_{d-1}(y_2; \mu_2, \Omega_{22}, \nu),$$

namely it is a $(d-1)$-dimensional central t pdf. The joint density of Y is equal to

$$
\begin{aligned}
&f_{Y_2}(y_2) f_{Y_1|Y_2=y_2}(y_1) \\
&= \psi_{d-1}(y_2; \mu_2, \Omega_{22}, \nu) \int_0^\infty \frac{v^{(\nu+d-1)/2-1} e^{-v}}{\Gamma(\frac{\nu+d-1}{2})} \phi\left\{ (\Omega_{1\cdot 2})^{-1/2} (y_1 - \mu_{1\cdot 2}) \sqrt{\frac{2v}{\nu+d-1}} - (\Omega_{11\cdot 2})^{-1/2} \kappa_1 \right\} \mathrm{d}v \\
&= \int_0^\infty \frac{(\Omega_{11\cdot 2})^{-1/2} v^{\nu/2-1} e^{-v}}{\Gamma(\frac{\nu}{2})} \left(\frac{2v}{\nu}\right)^{d/2} \phi_{d-1}\left(\frac{y_2 - \mu_2}{\sqrt{\frac{\nu}{2v}}} \right) \phi\left\{ (\Omega_{11\cdot 2})^{1/2} (y_1 - \mu_{1\cdot 2}) \sqrt{\frac{2v}{\nu}} - \kappa_1 \right\} \mathrm{d}v \\
&= \int_0^\infty \frac{v^{\nu/2-1} e^{-v}}{\Gamma(\frac{\nu}{2})} \phi_d \left\{ \begin{pmatrix} y_1 - \mu_1 - \kappa_1 \sqrt{\frac{\nu}{2v}} \\ y_2 - \mu_2 \end{pmatrix}; \sqrt{\frac{\nu}{2v}} \Omega \right\} \mathrm{d}v
\end{aligned}
$$

□

D.1.2 Proof of Proposition 4.5

Let $Z_s(h) = Z(s+h) - Z(s)$, with $s, h \in \mathbb{R}$. From Azzalini (e.g. 2013, Ch. 5) it follows that $Z_s(h) \sim \mathcal{SN}(\omega_s^2(h), \alpha_s(h), \tau)$, where

$$\omega_s^2(h) = 2(1 - \delta(s+h)\delta(s) - \rho(h)[\{1-\delta^2(s+h)\}\{1-\delta^2(s)\}]^{1/2})$$

and

$$\alpha_s^2(h) = \frac{\delta(s+h) - \delta(s)}{\{2(1 - \{\delta(s+h) + \delta(s)\}/2 - \rho(h)[\{1-\delta^2(s+h)\}\{1-\delta^2(s)\}]^{1/2})\}^{1/2}}.$$

Therefore, it follows that $\mathrm{Var}\{Z_s(h)\} = 2\{1 - r_s(h)\} \equiv \sigma_s^2(h)$, where

$$r_s(h) = r/2\{\delta(s+h) - \delta(s)\}^2 + \delta(s+h)\delta(s) + \rho(h)[\{1-\delta^2(s+h)\}\{1-\delta^2(s)\}]^{1/2}$$

and r is as in (4.9). By the assumption on $\delta(s)$ we have that $r_s(h) = \rho(h) + \delta^2(s)(1-\rho(h)) + o(1)$, as $h \to 0$. Using this and the assumption on $\rho(h)$ we obtain $1 - r_s(h) = O(|\log|h||^{-a})$, as $h \to 0$. Now, if $Z \sim \mathcal{SN}(\alpha, \tau)$, then $\Pr(|Z| > z) \leq 2\{1 - \Phi(z; |\alpha|, \tau)\}$. Furthermore, for $\alpha > 0$ the tail behavior of the skew-normal distribution satisfies the condition

$$1 - \Phi(z; \alpha, \tau) \leq K^{-1}\{1 - \Phi(z)\}, \quad x > 0,$$

where $K = \Phi\{\tau/(1+\alpha^2)^{1/2}\}$. Putting these two results together we obtain that for a non-decreasing function $g(h)$, the following inequality

$$\Pr\{|Z_s(h)| > g(h)\} \leq 2K^{-1}[1 - \Phi\{g(h)/\sigma_s(h)\}],$$

holds. Note that we obtained then the same inequality of Lindgren (2012, page 48) apart from the constant K. Therefore, the remaining part of the proof follows in the same way as the proof provided therein.

D.1.3 Proof of Proposition 4.6

Recall that if $Z \sim \mathcal{SN}_2(\bar{\Omega}, \alpha)$, then $Z_j \sim \mathcal{SN}(\alpha_j^*)$ and $Z_j|Z_{3-j} \sim \mathcal{SN}(\alpha_{j\cdot 3-j})$ for j=1,2 (e.g. Azzalini (2013, Ch. 2) or Proposition 4.2), where

$$\alpha_j^* = \frac{\alpha_j + \omega\alpha_{3-j}}{\sqrt{1+\alpha_{3-j}^2(1-\omega^2)}}, \quad \alpha_{j\cdot 3-j} = \alpha_j\sqrt{1-\omega^2}.$$

Define $x_j(u) = \Phi^{\leftarrow}(1-u;\alpha_j^*)$, for any $u \in [0,1]$, where $\Phi^{\leftarrow}(\cdot;\alpha_j^*)$ is the inverse of the marginal distribution function $\Phi(\cdot;\alpha_j^*)$, $j=1,2$. The asymptotic behaviour of $x_j(u)$ as $u \to 0$ is

$$x_j(u) = \begin{cases} x(u), & \text{if } \alpha_j^* \geq 0 \\ x(u)/\bar{\alpha}_j - \{2\log(1/u)\}^{-1/2}\log(\sqrt{\pi}\alpha_j^*), & \text{if } \alpha_j^* < 0 \end{cases} \tag{D.1}$$

for $j=1,2$, where $\bar{\alpha}_j = \{1+\alpha_j^{*2}\}^{1/2}$ and $x(u) \approx \{2\log(1/u)\}^{1/2} - \{2\log(1/u)\}^{-1/2}\{\log\log(1/u) + \log(2\sqrt{\pi})\}$ (Padoan, 2011). The limiting behaviour of the joint survivor function of the bivariate skew-normal distribution is described by

$$p(u) = \Pr\{Z_1 > x_1(u), Z_2 > x_2(u)\}, \qquad u \to 0. \tag{D.2}$$

For case (a), when $\alpha_1, \alpha_2 > 0$, then $x_1(u) = x_2(u) = x(u)$, and the joint upper tail (D.2) behaves as

$$\begin{aligned} p(u) &= \int_{x(u)}^{\infty} \left\{1 - \Phi\left(\frac{y(u)-\omega v}{\sqrt{1-\omega^2}}; \alpha_{1\cdot 2}\right)\right\} \phi(v;\alpha_2^*)\mathrm{d}v \\ &\approx \frac{\sqrt{1-\omega^2}}{x(u)} \int_0^{\infty} \frac{\phi_2(x(u), x(u)+t/x(u); \bar{\Omega}, \alpha)}{x(u)(1-\omega) - \omega t/x(u)}\mathrm{d}t \\ &\approx \frac{e^{-x^2(u)/(1+\omega)}}{\pi(1-\omega)x^2(u)} \left(\int_0^{\infty} e^{-t/(1+\omega)}\mathrm{d}t - \frac{e^{-x^2(u)(\alpha_1+\alpha_2)^2/2}}{\sqrt{2\pi}(\alpha_1+\alpha_2)x(u)} \int_0^{\infty} e^{-t\{1/(1+\omega)+\alpha_2(\alpha_1+\alpha_2)\}}\mathrm{d}t\right) \\ &= \frac{e^{-x^2(u)/(1+\omega)}(1+\omega)}{\pi(1-\omega)x(u)^2} \left(1 - \frac{e^{-x^2(u)(\alpha_1+\alpha_2)^2/2}}{\sqrt{2\pi}(\alpha_1+\alpha_2)\{1+\alpha_2(\alpha_1+\alpha_2)(1+\omega)\}x(u)}\right), \end{aligned} \tag{D.3}$$

as $u \to 0$. The first approximation is obtained by using $1 - \Phi(x;\alpha) \approx \phi(x;\alpha)/x$ as $x \to +\infty$, when $\alpha > 0$ (Padoan, 2011). The second approximation uses $1 - \Phi(x) \approx \phi(x)/x$ as $x \to +\infty$ (Feller, 1968). Let $X_j = \{-1/\log\Phi(Z_j;\alpha_j^*)\}$, $j=1,2$. Substituting $x(u)$ into (D.3) substituting and using the approximation $1 - \Pr(X_j > x) \approx 1/x$ as $x \to \infty$, $j=1,2$, we obtain that (D.2) with common unit Fréchet margins behaves asymptotically as $\mathscr{L}(x)\ x^{-2/(1+\omega)}$, as $x \to +\infty$, where

$$\mathscr{L}(x) = \frac{2(1+\omega)(4\pi\log x)^{-\omega/(1+\omega)}}{1-\omega} \left(1 - \frac{(4\pi\log x)^{\{(\alpha_1+\alpha_2)^2-1\}/2}\, x^{-(\alpha_1+\alpha_2)^2}}{(\alpha_1+\alpha_2)\{1+\alpha_2(\alpha_1+\alpha_2)(1+\omega)\}}\right). \tag{D.4}$$

As the second term in the parentheses in (D.4) is $o(x^{(\alpha_1+\alpha_2)})$, then the quantity inside the parentheses $\to 1$ rapidly as $x \to \infty$, and so $\mathscr{L}(x)$ is well approximated by the first term in (D.4). When $\alpha_2 < 0$ and $\alpha_1 \geq -\alpha_2/\omega$, then $\alpha_1^*, \alpha_2^* > 0$ and we obtain the same outcome.

For case (b), when $\alpha_2 < 0$ and $-\omega, \alpha_2 \leq \alpha_1 < -\omega^{-1}\alpha_2$, then $\alpha_1^* \geq 0$ and $\alpha_2^* < 0$ and hence $x_1(u) = x(u)$ and $x_2(u) \approx x(u)/\bar{\alpha}_2$ as $u \to 0$. When $\alpha_1 > -\bar{\alpha}_2\alpha_2$, then following a similar derivation to those in (D.3), we obtain that

$$p(u) \approx \frac{\bar{\alpha}_2^2(1-\omega^2)(1-\omega\bar{\alpha}_2)^{-1}}{\pi(\bar{\alpha}_2-\omega)x^2(u)} \exp\left[-\frac{x^2(u)}{2}\left\{\frac{1-\omega^2+(\bar{\alpha}_2-\omega)^2}{(1-\omega^2)\bar{\alpha}_2^2}\right\}\right], \quad \text{as } u \to 0.$$

Similarly, when $\alpha_1 < -\bar{\alpha}_2\alpha_2$, and noting that $\Phi(x) \approx -\phi(-x)/x$ as $x \to -\infty$, then

$$p(u) \approx \frac{-\bar{\alpha}_2^2\{1-\omega\bar{\alpha}_2+\alpha_2(\alpha_2+\alpha_1\bar{\alpha}_2)(1-\omega^2)\}^{-1}}{\pi(\bar{\alpha}_2-\omega)(1-\omega^2)^{-1}(\alpha_1+\alpha_2/\bar{\alpha}_2)x^3(u)} e^{-\frac{x^2(u)}{2}\left\{\frac{1-\omega^2+(\bar{\alpha}_2-\omega)^2}{(1-\omega^2)\bar{\alpha}_2^2}+\left(\alpha_1+\frac{\alpha_2}{\bar{\alpha}_2}\right)^2\right\}}, \quad \text{as } u \to 0.$$

For case (c), when $\alpha_2 < 0$ and $0 < \alpha_1 < -\omega\alpha_2$, then $\alpha_1^*, \alpha_2^* < 0$ and hence $x_1(u) \approx x(u)/\bar{\alpha}_1$ and $x_2(u) \approx x(u)/\bar{\alpha}_2$ as $u \to 0$. Then as $u \to 0$ we have

$$\begin{aligned} p(u) \approx{}& \frac{-\bar{\alpha}_2^{3/2}\bar{\alpha}_1^2(1-\omega^2)(\bar{\alpha}_2-\omega\bar{\alpha}_1)^{-1}(\alpha_1\bar{\alpha}_2+\alpha_2\bar{\alpha}_1)^{-1}}{\pi\{1-\omega\bar{\alpha}_2+\alpha_2(\alpha_2+\alpha_1\bar{\alpha}_2/\bar{\alpha}_1)(1-\omega^2)\}x^3(u)} \\ &\times \exp\left[-\frac{x^2(u)}{2(1-\omega^2)}\left(\frac{\alpha_1^2(1-\omega^2)+1}{\bar{\alpha}_1^2}+\frac{\alpha_2^2(1-\omega^2)+1}{\bar{\alpha}_2^2}+\frac{2(\alpha_1\alpha_2(1-\omega^2)-\omega)}{\bar{\alpha}_1\bar{\alpha}_2}\right)\right] \quad u \to 0. \end{aligned}$$

When $\alpha_1, \alpha_2 < 0$ and $\omega_2^{-1}\alpha_2 \leq \alpha_1 < 0$ the same argument holds. Finally, interchanging α_1 with α_2 produces the same results but substituting α_j and $\bar{\alpha}_j$ with α_{3-j} and $\bar{\alpha}_{3-j}$ respectively, for $j = 1, 2$.

D.1.4 Proof of Proposition 4.3

Let $Z \sim \mathcal{ST}(\alpha, \tau, \kappa, \nu)$. Then $1 - \Psi(x; \alpha, \tau, \nu) \approx x^{-\nu}\mathscr{L}(x; \alpha, \tau, \nu)$ as $x \to +\infty$, for any $\nu > 1$, where

$$\mathscr{L}(x; \alpha, \tau, \kappa, \nu) = \frac{\Gamma\{(\nu+1)/2\}\Psi(\alpha\sqrt{\nu+1}; \nu+1)}{\Gamma(\nu/2)\sqrt{\pi}\nu^{3/2}\Psi(\tau/\sqrt{1+\alpha^2}; \kappa/\sqrt{1+\alpha^2}, \nu)}\left(\frac{1}{x^2}+\frac{1}{\nu}\right)^{-(\nu+1)/2}$$

is a slowly varying function (e.g de Haan and Ferreira, 2006, Appendix B). From Corollary 1.2.4 in de Haan and Ferreira (2006), it follows that the normalisation constants are $a_n = \Psi^{\leftarrow}(1 - 1/n; \alpha, \tau, \kappa, \nu)$, where $\Psi^{\leftarrow}$ is the inverse function of Ψ, and $b_n = 0$, and therefore $a_n = \{n\mathscr{L}(\alpha, \tau, \kappa, \nu)\}^{1/\nu}$, where $\mathscr{L}(\alpha, \tau, \kappa, \nu) \equiv \mathscr{L}(\infty; \alpha, \tau, \kappa, \nu)$. Applying Theorem 1.2.1 in de Haan and Ferreira (2006) we obtain that $M_n/a_n \Rightarrow U$, where U has ν-Fréchet univariate marginal distributions.

Let $Z \sim \mathcal{ST}_d(\bar{\Omega}, \alpha, \tau, \kappa, \nu)$. For any $j \in \{1, \ldots, d\}$ consider the partition $Z = (Z_j, Z_{I_j}^\top)^\top$, where $I_j = \{1, \ldots, d\}\backslash j$ and $Z_j = Z_{\{j\}}$, and the respective partition of $(\bar{\Omega}, \alpha)$. Define $a_n = (a_{n,1}, \ldots, a_{n,d})$, where $a_{n,j} = \{n\mathscr{L}(\alpha_j^*, \tau_j^*, \kappa_j^*, \nu)\}^{1/\nu}$ and $\alpha_j^* = \alpha_{\{j\}}^*$, $\tau_j^* = \tau_{\{j\}}^*$ and $\kappa_j^* = \kappa_{\{j\}}^*$ are the marginal parameters (4.5) under Proposition 4.2(1). From Theorem 6.1.1 and Corollary 6.1.3 in de Haan and Ferreira (2006), $M_n/a_n \Rightarrow U$, where the distribution of U is $G(x) = \exp\{-V(x)\}$ with $V(x) = \lim_{n\to+\infty} n\{1 - \Pr(Z_1 \leq a_{n,1}x_1, \ldots, Z_d \leq a_{n,d}x_d)\}$ for all $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}_+^d$.

Applying the conditional tail dependence function framework of Nikoloulopoulos et al. (2009) it follows that

$$V(x_j, i \in I) = \lim_{n\to\infty} \sum_{j=1}^{d} x_j^{-\nu} \Pr(Z_i \leq a_{n,i} x_i, i \in I_j | Z_j = a_{n,j} x_j).$$

From the conditional distribution in Proposition 4.2(1) we have that

$$\left\{ \left(\frac{Z_i - a_{n,j} x_j}{\{\zeta_{n,j}(1-\omega_{i,j}^2)\}^{1/2}}, i \in I_j \right)^\top | Z_j = a_{n,j} x_j \right\} \sim \mathcal{ST}_{d-1}\left(\bar{\Omega}_j^+, \alpha_j^+, \tau_{n,j}, \kappa_{n,j}, \nu+1\right),$$

for $j \in \ldots 1, \ldots, d$, where $\bar{\Omega}_j^+ = \omega_{I_jI_j\cdot j}^{-1} \Omega_{I_jI_j\cdot j} \omega_{I_jI_j\cdot j}^{-1}$, $\omega_{I_jI_j\cdot j} = \mathrm{diag}(\Omega_{I_jI_j\cdot j})^{1/2}$, $\bar{\Omega}_{I_jI_j\cdot j} = \bar{\Omega}_{I_jI_j} - \bar{\Omega}_{I_jj}\bar{\Omega}_{jI_j}$, $\alpha_j^+ = \bar{\Omega}_{I_jI_j\cdot j}\alpha_{I_j}$ $\zeta_{n,j} = [\nu + (a_{n,j}x_j)^2]/(\nu+1)$, $\tau_{n,j} = [(\bar{\Omega}_{jI_j}\alpha_{I_j} + \alpha_j)a_{n,j}x_j + \tau]/\zeta_{n,j}^{1/2}$ and $\kappa_{n,j} = \kappa/\zeta_{n,j}^{1/2}$. Now, for any $j \in \{1, \ldots, d\}$ and all $i \in I_j$

$$\frac{a_{n,i}x_i - a_{n,j}x_j}{\{\zeta_{n,j}(1-\omega_{i,j}^2)\}^{1/2}} \to \frac{(x_i^*/x_j^* - \omega_{i,j})(\nu+1)^{1/2}}{\{(1-\omega_{i,j})\}^{1/2}} \quad \text{as } n \to +\infty,$$

where $\omega_{i,j}$ is the (i,j)-th element of $\bar{\Omega}$, $x_j^+ = x_j \mathscr{L}^{1/\nu}(\alpha_j^*, \tau_j^*, \kappa_j^*, \nu)$ and $\tau_{n,j} \to \tau_j^+ = (\bar{\Omega}_{jI_j}\alpha_{I_j} + \alpha_j)(\nu+1)^{1/2}$, and $\kappa_{n,j} \to 0$ as $n \to +\infty$. As a consequence

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu} \Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}} \left(\frac{x_i^+}{x_j^+} - \omega_{i,j} \right), i \in I_j \right)^\top ; \bar{\Omega}_j^+, \alpha_j^+, \tau_j^+, \nu+1 \right).$$

D.1.5 Proof of Theorem 4.7

Let $Y(s)$ be a skew-normal process with finite dimensional distribution $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$. For any $j \in I = \{1, \ldots, d\}$ consider the partition $Y = (Y_j, Y_{I_j}^\top)^\top$, where $I_j = I \backslash j$, $Y_j = Y_{\{j\}} = Y(s_j)$ and $Y_{I_j} = (Y_i, i \in I_j)^\top$, and the respective partition of $(\bar{\Omega}, \alpha)$. The exponent function (4.14) is

$$V(x_j, j \in I) = \mathbb{E}\left[\max_j \left\{ \frac{(Y_j^+/x_j)^\xi}{m_j^+} \right\} \right] = \int_{\mathbb{R}^d} \max_j \left\{ \frac{(y_j/x_j)^\xi}{m_j^+}, 0 \right\} \phi_d(y; \bar{\Omega}; \alpha, \tau) \mathrm{d}y,$$

where $x_j \equiv x(s_j)$, $y_j \equiv y(s_j)$ and $m_j^+ \equiv m^+(s_j)$. Then

$$V(x_j, j \in I) = \sum_{j=1}^{d} V_j, \quad V_j = \frac{1}{m_j^+} \int_0^\infty \left(\frac{y_j}{x_j} \right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d(y; \bar{\Omega}; \alpha, \tau) \mathrm{d}y_{I_j} \mathrm{d}y_j, \tag{D.5}$$

where $x_{I_j} = (x_i, i \in I_j)^\top$ and $y_{I_j} = (y_i, i \in I_j)^\top$. As $Y_j \sim \mathcal{SN}(\alpha_j^*, \tau_j^*)$, where $\alpha_j^* = \alpha_{\{j\}}^*$ and $\tau_j^* = \tau_{\{j\}}^*$ are the marginal parameters derived from Proposition 4.2(1), then

$$\begin{aligned} m_j^+ = \int_0^\infty y_j^\nu \, \phi(y_j; \alpha_j^*, \tau_j^*) \mathrm{d}y_j &= \frac{1}{\Phi\{\tau_j^*(1+\alpha_j^{*2})^{-1/2}\}} \int_0^\infty y_j^\nu \, \phi(y_j) \Phi(\alpha_j^* y_j + \tau_j^*) \mathrm{d}y_j \\ &= \frac{2^{(\nu-2)/2}\Gamma\{(\nu+1)/2\}\Psi(\alpha_j^*\sqrt{\nu+1}; -\tau_j^*, \nu+1)}{\sqrt{\pi}\Phi[\tau\{1+Q_{\bar{\Omega}}(\alpha)\}^{-1/2}]} \end{aligned}$$

by observing that $\tau_j^*\{1+\alpha_j^{*2}\}^{1/2} = \tau\{1+Q_{\bar{\Omega}}(\alpha)\}^{-1/2}$.

For $j = 1, \ldots, d$ define $x_j^\circ = x_j(m_j^+)^{1/\nu}$ and $m_j^+ = \bar{m}_j^+/\Phi[\tau\{1+Q_{\bar{\Omega}}(\alpha)\}^{-1/2}]$, where $\bar{m}_j^+ = (\pi)^{1/2}2^{(\nu-2)/2}\Gamma\{(\nu+1)/2\}\Psi(\alpha_j^*\sqrt{\nu+1}; -\tau_j^*, \nu+1)$. Then, for any $j = 1, \ldots, d$

$$\begin{aligned}
V_j &= \frac{1}{m_j^+}\int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d(y;\bar{\Omega},\alpha,\tau)\mathrm{d}y_{I_j}\mathrm{d}y_j \\
&= \frac{1}{\bar{m}_j^+}\int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d(y;\Omega)\Phi(\alpha^\top y+\tau)\mathrm{d}y_{I_j}\mathrm{d}y_j \\
&= \frac{1}{\bar{m}_j^+}\int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \phi(y_j)\int_{-\infty}^{y_j x_{I_j}/x_j} \phi_{d-1}(y_{I_j} - y_j\bar{\Omega}_{j,I_j};\bar{\Omega}_j^\circ)\Phi(\alpha^\top y+\tau)\mathrm{d}y_{I_j}\mathrm{d}y_j \\
&= \frac{1}{\bar{m}_j^+}\int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \phi(y_j)\Phi_d\left(y_j^\circ;\Omega_j^{\circ\circ}\right)\mathrm{d}y_j,
\end{aligned}$$

where

$$y_j^\circ = \begin{pmatrix} y_j\,\omega_{I_jI_j\cdot j}^{-1}(x_{I_j}^\circ/x_j^\circ - \bar{\Omega}_{I_jj}) \\ y_j\alpha_j^* + \tau_j^* \end{pmatrix},$$

with $\omega_{I_jI_j\cdot j} = \mathrm{diag}(\bar{\Omega}_{I_jI_j\cdot j})^{1/2}$, $\bar{\Omega}_{I_jI_j\cdot j} = \bar{\Omega}_{I_jI_j} - \bar{\Omega}_{I_jj}\bar{\Omega}_{jI_j}$, $y_j\alpha_j^* + \tau_j^* = \frac{y_j(\alpha_j+\bar{\Omega}_{jj}^{-1}\bar{\Omega}_{jI_j}\alpha_{I_j})+\tau}{\{1+Q_{\bar{\Omega}_{I_jI_j\cdot j}}(\alpha_{I_j})\}^{1/2}}$ and

$$\Omega_j^{\circ\circ} = \begin{pmatrix} \bar{\Omega}_j^\circ & -\frac{\bar{\Omega}_{I_jI_j\cdot j}\omega_{I_jI_j\cdot j}^{-1}\alpha_{I_j}}{\{1+Q_{\bar{\Omega}_{I_jI_j\cdot j}}(\alpha_{I_j})\}^{1/2}} \\ -\left(\frac{\bar{\Omega}_{I_jI_j\cdot j}\omega_{I_jI_j\cdot j}^{-1}\alpha_{I_j}}{\{1+Q_{\bar{\Omega}_{I_jI_j\cdot j}}(\alpha_{I_j})\}^{1/2}}\right)^\top & 1 \end{pmatrix},$$

where $\bar{\Omega}_j^\circ = \omega_{I_jI_j\cdot j}^{-1}\,\bar{\Omega}_{I_jI_j\cdot j}\,\omega_{I_jI_j\cdot j}^{-1}$ and $\frac{\Omega_{I_jI_j\cdot j}\omega_{I_jI_j\cdot j}^{-1}\alpha_{I_j}}{\{1+Q_{\Omega_{I_jI_j\cdot j}}(\alpha_{I_j})\}^{1/2}} = \frac{\Omega_j^\circ\,\omega_{I_jI_j\cdot j}\,\alpha_{I_j}}{\{1+Q_{\bar{\Omega}_j^\circ}(\omega_{I_jI_j\cdot j}\alpha_{I_j})\}^{1/2}}$.

Applying Dutt's (Dutt, 1973) probability integrals we obtain

$$\begin{aligned}
V_j &= \frac{1}{\bar{m}_j^+}\int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \phi(y_j)\Phi_d\left(y_j^\circ;\Omega_j^{\circ\circ}\right)\mathrm{d}y_j, \\
&= \frac{1}{x_j^\nu}\frac{\Psi_{d+1}\left(\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^\circ}{x_j^\circ}-\omega_{i,j}\right), i\in I_j\right), \alpha_j^*\sqrt{\nu+1}\right)^\top; \Omega_j^{\circ\circ}, \left(0, -\tau_j^*\right)^\top, \nu+1\right)}{\Psi(\alpha_j^*\sqrt{\nu+1}; -\tau_j^*, \nu+1)}.
\end{aligned}$$

This is recognised as the form of a $(d-1)$-dimensional non-central extended skew-t distribution with $\nu+1$ degrees of freedom (Jamalizadeh et al., 2009), from which V_j can be expressed as

$$V_j = \frac{1}{x_j^\nu}\Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^\circ}{x_j^\circ}-\omega_{i,j}\right), i\in I_j\right)^\top; \bar{\Omega}_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, \nu+1\right)$$

for $j = 1, \ldots, d$ where $\alpha_j^\circ = \omega_{I_jI_j\cdot j}\,\alpha_{I_j}$, $\tau_j^\circ = (\bar{\Omega}_{jI_j}\alpha_{I_j} + \alpha_j)(\nu+1)^{1/2}$ and $\kappa_j^\circ = -\{1 + Q_{\bar{\Omega}_{I_jI_j\cdot j}}(\alpha_{I_j})\}^{-1/2}\tau$. Substituting the expression for V_j into (D.5) then gives the required the

exponent function.

D.2 Computational details

D.2.1 Computation of d-dimensional extremal-skew-t density for $d = 2, 3, 4$.

For clarity of exposition we focus on the finite dimensional distribution G of the extremal-t process. We initially assume that $\alpha = 0$ and $\tau = 0$, and relax this assumption later. For brevity the exponent function is written as

$$V(x_j, j \in I) = \sum_{j \in I} x_j^{-1} T_j, \quad T_j = \Psi_{d-1}\left(u_j; \bar{\Omega}_j^{\circ}, \nu + 1\right)$$

where $I = \{1, \ldots, d\}$, $u_j = \left[\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left\{\left(\frac{x_i}{x_j}\right)^{1/\nu} - \omega_{i,j}\right\}, i \in I_j\right]^{\top}$ and where $I_j = I \backslash \{j\}$. By successive differentiations the 2-dimensional density ($d = 2$) is

$$f(x) = (-V_{12} + V_1 V_2) G(x), \quad x \in \mathbb{R}_+^2,$$

the 3-dimensional density ($d = 3$) is

$$f(x) = (-V_{123} + V_1 V_{23} + V_2 V_{13} + V_3 V_{12} - V_1 V_2 V_3) G(x), \quad x \in \mathbb{R}_+^3$$

and the 4-dimensional density ($d = 4$) is

$$\begin{aligned} f(x) = (&-V_{1234} + V_1 V_{234} + V_2 V_{134} + V_3 V_{124} + V_{12} V_{34} + V_{13} V_{24} + V_{14} V_{23} \\ &- V_1 V_2 V_{34} - V_1 V_3 V_{24} - V_1 V_4 V_{23} - V_2 V_3 V_{14} - V_2 V_4 V_{13} - V_3 V_4 V_{12} \\ &+ V_1 V_2 V_3 V_4) G(x), \quad x \in \mathbb{R}_+^4 \end{aligned}$$

where $V_{i_1,\ldots,i_m} := \frac{\mathrm{d}^m V(x_j, j \in I)}{\mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_m}}$ for $i_k \in I$. The derivatives of the exponent function are given by

$$V_{i_1,\ldots,i_m} = \sum_{k=1}^{d} x_{i_k}^{-1} \frac{\mathrm{d}^m T_{i_k}}{\mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_m}} - \sum_{\ell=1}^{m} x_{i_\ell}^{-2} \frac{\mathrm{d}^{m-1} T_{i_\ell}}{\mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_{\ell-1}} \mathrm{d}x_{i_{\ell+1}} \cdots \mathrm{d}x_{i_m}}. \tag{D.6}$$

In particular, when $m = d$ it follows that $\{i_1, \ldots, i_d\} = \{1, \ldots, d\}$ and that

$$V_{1\cdots d} = -(\nu x_1)^{-(d+1)} \psi_{d-1}\left(u_1; \bar{\Omega}_1^{\circ}, \nu + 1\right) \prod_{i=2}^{d} \sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}} \left(\frac{x_i}{x_1}\right)^{\frac{1}{\nu}-1}.$$

When $d = 2$ or 3, the derivatives of T_j, for $j \in I$ are given by

$$\frac{\mathrm{d}T_j}{\mathrm{d}x_{i_1}} = \sum_{p=1}^{d-1} \frac{\mathrm{d}}{\mathrm{d}u_{p,j}} \Psi_{d-1}\left(u_j; \bar{\Omega}_j^{\circ}, \nu + 1\right) \frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_1}}, \tag{D.7}$$

$$\frac{\mathrm{d}^2 T_j}{\mathrm{d}x_{i_1}\mathrm{d}x_{i_2}} = \sum_{p=1}^{d-1}\left(\frac{\mathrm{d}}{\mathrm{d}u_{p,j}}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\frac{\mathrm{d}^2 u_{p,j}}{\mathrm{d}x_{i_1}\mathrm{d}x_{i_2}} + \frac{\mathrm{d}^2}{\mathrm{d}u_{p,j}^2}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_1}}\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_2}}\right) + \sum_{p=1}^{d-2}\sum_{q=p+1}^{d-1}\frac{\mathrm{d}^2}{\mathrm{d}u_{p,j}\mathrm{d}u_{q,j}}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\left[\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_1}}\frac{\mathrm{d}u_{q,j}}{\mathrm{d}x_{i_2}} + \frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_2}}\frac{\mathrm{d}u_{q,j}}{\mathrm{d}x_{i_1}}\right]. \tag{D.8}$$

where $u_{p,j}$ is the p-th element of u_j, and when $d = 3$

$$\begin{aligned}\frac{\mathrm{d}^3 T_j}{\mathrm{d}x_{i_1}\mathrm{d}x_{i_2}\mathrm{d}x_{i_3}} &= \sum_{p=1}^{2}\sum_{q=2}^{3}\left(\frac{\mathrm{d}^2}{\mathrm{d}u_{p,j}\mathrm{d}u_{q,j}}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\sum_{\substack{r,s,t\in I\\ r\neq s\neq t}}\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_r}}\frac{\mathrm{d}^2 u_{q,j}}{\mathrm{d}x_{i_s}\mathrm{d}x_{i_t}} + \frac{\mathrm{d}u_{q,j}}{\mathrm{d}x_{i_r}}\frac{\mathrm{d}^2 u_{p,j}}{\mathrm{d}x_{i_s}\mathrm{d}x_{i_t}}\right)\\ &+ \sum_{p=1}^{3}\sum_{\substack{q=1\\ q\neq p}}^{3}\frac{\mathrm{d}^3}{\mathrm{d}u_{p,j}^2\mathrm{d}u_{q,j}}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\sum_{\substack{r,s,t\in I\\ r\neq s\neq t}}\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_r}}\frac{\mathrm{d}u_{p,j}}{\mathrm{d}x_{i_s}}\frac{\mathrm{d}u_{q,j}}{\mathrm{d}x_{i_t}}\\ &+ \frac{\mathrm{d}^3}{\mathrm{d}u_{1,j}\mathrm{d}u_{2,j}\mathrm{d}u_{3,j}}\Psi_{d-1}\left(u_j;\bar{\Omega}_j^{\circ},\nu+1\right)\sum_{\substack{r,s,t\in I\\ r\neq s\neq t}}\frac{\mathrm{d}u_{1,j}}{\mathrm{d}x_{i_r}}\frac{\mathrm{d}u_{2,j}}{\mathrm{d}x_{i_s}}\frac{\mathrm{d}u_{3,j}}{\mathrm{d}x_{i_t}}.\end{aligned} \tag{D.9}$$

We provide the derivatives of the d-dimensional t cdf below. When $d = 1$ and for all $x \in \mathbb{R}_+$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}\Psi(x;\nu) = \psi(x;\nu), \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}\Psi(x;\nu) = -\frac{(\nu+1)x}{\nu+x^2}\psi(x;\nu),\\ &\frac{\mathrm{d}^3}{\mathrm{d}x^3}\Psi(x;\nu) = \frac{(\nu+1)(x^2-\nu+(\nu+1)x^2)}{(\nu+x^2)^2}\psi(x;\nu).\end{aligned}$$

When d=2 and for all $x \in \mathbb{R}_+^2$,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x_1}\Psi_2(x;\bar{\Omega},\nu) = \psi(x_1;\nu)\Psi\left(v_{2\cdot 1};\nu+1\right),\\ &\frac{\mathrm{d}^2}{\mathrm{d}x_1^2}\Psi_2(x;\bar{\Omega},\nu) = -\psi(x_1;\nu)\left\{\frac{(\nu+1)x_1}{\nu+x_1^2}\Psi\left(v_{2\cdot 1};\nu+1\right) + \sqrt{\frac{\nu+1}{1-\omega^2}}\left(\frac{\omega\nu+x_2x_1}{(\nu+x_1^2)^{3/2}}\right)\psi\left(v_{2\cdot 1};\nu+1\right)\right\},\\ &\frac{\mathrm{d}^2}{\mathrm{d}x_1\mathrm{d}x_2}\Psi_2(x;\bar{\Omega},\nu) = \psi_2(x;\bar{\Omega},\nu),\end{aligned}$$

where $v_{i\cdot j} = \sqrt{\frac{\nu+1}{\nu+x_j^2}}\frac{x_i-\omega_{i,j}x_1}{\sqrt{1-\omega_{i,j}^2}}, \quad j \in I, j \in I_j,$

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}x_1^3}\Psi_2(x;\bar{\Omega},\nu) &= \Psi\left(v_{2\cdot 1};\nu+1\right)\psi(x_1;\nu)\left\{\frac{(\nu+1)^2x_1^2-(\nu+1)(\nu-x_1^2)}{(\nu+x_1^2)^2}\right\}\\ &+ \psi\left(v_{2\cdot 1};\nu+1\right)\psi(x_1;\nu)\sqrt{\frac{\nu+1}{1-\omega^2}}\frac{1}{(\nu+x_1^2)^{5/2}}\\ &\times\Big\{x_1(\omega\nu+x_2x_1)(2\nu-1) - x_2(\nu+x_1^2)\\ &- \frac{(\omega(\nu+x_1^2)+(x_2-\omega x_1)x_1)(\nu+2)(x_2-\omega x_1)(\omega\nu+x_2x_1)}{(\nu+x_1^2)(1-\omega^2)+(x_2-\omega x_1)^2}\Big\},\end{aligned}$$

$$\frac{\mathrm{d}^3}{\mathrm{d}x_1^2\mathrm{d}x_2}\Psi_2(x;\bar{\Omega},\nu) = -\frac{(\nu+2)(x_1-\omega x_2)}{2\pi\nu(1-\omega^2)^{3/2}}\left(1+\frac{x_1^2-2\omega x_1x_2+x_2^2}{\nu(1-\omega^2)}\right)^{-(\frac{\nu}{2}+1)}.$$

When d=3 and for all $x\in\mathbb{R}^3_+$,

$$\frac{\mathrm{d}}{\mathrm{d}x_1}\Psi_3(x;\bar{\Omega},\nu) = \psi(x;\nu)\Psi_2\left\{(v_{2\cdot1},v_{3\cdot1})^\top;\bar{\Omega}_1^\circ,\nu+1\right\},$$

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}x_1^2}\Psi_3(x;\bar{\Omega},\nu) = \frac{-\psi(x_1;\nu)}{\nu+x_1^2}\Big[&(\nu+1)x_1\times\Psi_2\left\{(v_{2\cdot1},v_{3\cdot1})^\top;\bar{\Omega}_1^\circ,\nu+1\right\}\\
&+\psi\left(v_{2\cdot1};\nu+1\right)\sqrt{\frac{\nu+1}{1-\omega_{12}^2}}\frac{x_2x_1+\omega_{12}\nu}{\sqrt{\nu+x_1^2}}\\
&\times\Psi\left(\frac{\sqrt{\nu+2}\left\{(x_3-\omega_{13}x_1)(1-\omega_{12}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_2-\omega_{12}x_1)\right\}}{\sqrt{\left\{(1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2\right\}\left\{(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right\}}};\nu+2\right)\\
&+\psi\left(v_{3\cdot1};\nu+1\right)\sqrt{\frac{\nu+1}{1-\omega_{13}^2}}\frac{x_3x_1+\omega_{13}\nu}{\sqrt{\nu+x_1^2}}\\
&\times\Psi\left(\frac{\sqrt{\nu+2}\left\{(x_2-\omega_{12}x_1)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_3-\omega_{13}x_1)\right\}}{\sqrt{\left\{(1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2\right\}\left\{(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right\}}};\nu+2\right)\Big]\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}x_1\mathrm{d}x_2}\Psi_3(x;\bar{\Omega},\nu) = \psi(x_2;\nu)\psi\left(v_{1\cdot2};\nu+1\right)\sqrt{\frac{\nu+1}{(1-\omega_{12}^2)(\nu+x_2^2)}}\\
\times\Psi\left(\frac{\sqrt{\nu+2}\left\{(x_3-\omega_{23}x_2)(1-\omega_{12}^2)-(\omega_{13}-\omega_{12}\omega_{23})(x_1-\omega_{12}x_2)\right\}}{\sqrt{\left\{(1-\omega_{12}^2)(\nu+x_1^2)+(x_1-\omega_{12}x_2)^2\right\}\left\{(1-\omega_{12}^2)(1-\omega_{23}^2)-(\omega_{13}-\omega_{12}\omega_{23})^2\right\}}};\nu+2\right)\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}x_1^2\mathrm{d}x_2}\Psi_3(x;\bar{\Omega},\nu) = -\psi(x_3;\nu)\psi\left(v_{1\cdot3};\nu+1\right)\sqrt{\frac{\nu+1}{(1-\omega_{13}^2)(\nu+x_3^2)}}\Big[&\frac{(\nu+2)(x_1-\omega_{12}x_2)}{(1-\omega_{12}^2)(\nu+x_2^2)+(x_1-\omega_{12}x_2)^2}\\
&\times\Psi\left(\frac{\sqrt{\nu+2}\left\{(x_3-\omega_{23}x_2)(1-\omega_{12}^2)-(\omega_{13}-\omega_{12}\omega_{23})(x_1-\omega_{12}x_2)\right\}}{\sqrt{\left\{(1-\omega_{12}^2)(\nu+x_1^2)+(x-\omega_{12}x_2)^2\right\}\left\{(1-\omega_{12}^2)(1-\omega_{23}^2)-(\omega_{13}-\omega_{12}\omega_{23})^2\right\}}};\nu+2\right)\\
&+\frac{\sqrt{\nu+2}(1-\omega_{12}^2)}{\sqrt{(1-\omega_{12}^2)(1-\omega_{23}^2)-(\omega_{13}-\omega_{12}\omega_{23})^2}}\frac{(\omega_{13}-\omega_{12}\omega_{23})-(x_1-\omega_{12}x_2)(x_3-\omega_{23}x_2)}{\left\{(1-\omega_{12}^2)(\nu+x_2^2)+(x_1-\omega_{12}x_2)^2\right\}^{3/2}}\\
&\times\psi\left(\frac{\sqrt{\nu+2}\left\{(x_3-\omega_{23}x_2)(1-\omega_{12}^2)-(\omega_{13}-\omega_{12}\omega_{23})(x_1-\omega_{12}x_2)\right\}}{\sqrt{\left\{(1-\omega_{12}^2)(\nu+x_1^2)+(x_1-\omega_{12}x_2)^2\right\}\left\{(1-\omega_{12}^2)(1-\omega_{23}^2)-(\omega_{13}-\omega_{12}\omega_{23})^2\right\}}};\nu+2\right)\Big]\end{aligned}$$

$$
\begin{aligned}
\frac{\mathrm{d}^3}{\mathrm{d}x_1^3}\Psi_3(x;\bar{\Omega},\nu) = &-\frac{\psi(x_1;\nu)}{(\nu+x_1^2)}\left[\left(\frac{\nu+3}{\nu+x_1^2}\right)(1-x_1^2)(\nu+1)\Psi_2\left\{(v_{2\cdot1},v_{3\cdot1})^\top;\bar{\Omega}_1^\circ,\nu+1\right\}\right.\\
&+\Psi\left(\frac{\sqrt{\nu+2}\left[(x_3-\omega_{13}x_1)(1-\omega_{12}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_2-\omega_{12}x_1)\right]}{\sqrt{\left[(1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2\right]\left[(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right]}};\nu+2\right)\\
&\times\psi\left(v_{2\cdot1};\nu+1\right)\sqrt{\frac{\nu+1}{1-\omega_{12}^2}}\frac{2(x_2x_1+\omega_{12}\nu)(\nu+2)x_1-\nu(x_2-\omega_{12}x_1)}{(\nu+x_1^2)^{3/2}}\\
&\times\frac{(\nu+2)(x_2-\omega_{12}x_1)\sqrt{\nu+1}(x_2x_1+\omega_{12}\nu)^2}{\sqrt{1-\omega_{12}^2}(\nu+x_1^2)^{3/2}\left((1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2\right)}\\
&+\Psi\left(\frac{\sqrt{\nu+2}\left[(x_2-\omega_{12}x_1)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_3-\omega_{13}x_1)\right]}{\sqrt{\left[(1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2\right]\left[(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right]}};\nu+2\right)\\
&\times\psi\left(v_{3\cdot1};\nu+1\right)\sqrt{\frac{\nu+1}{1-\omega_{13}^2}}\frac{2(x_3x_1+\omega_{13}\nu)(\nu+2)x_1-\nu(x_3-\omega_{13}x_1)}{(\nu+x_1^2)^{3/2}}\\
&\times\frac{(\nu+2)(x_3-\omega_{13}x_1)\sqrt{\nu+1}(x_3x_1+\omega_{13}\nu)^2}{\sqrt{1-\omega_{13}^2}(\nu+x_1^2)^{3/2}\left((1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2\right)}\hat{\mathrm{E}}\\
&+\psi\left(\frac{\sqrt{\nu+2}\left[(x_3-\omega_{13}x_1)(1-\omega_{12}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_2-\omega_{12}x_1)\right]}{\sqrt{\left[(1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2\right]\left[(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right]}};\nu+2\right)\\
&\times\psi\left(v_{2\cdot1};\nu+1\right)\sqrt{\frac{(1-\omega_{12}^2)(\nu+2)}{(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2}}\\
&\times\frac{\sqrt{\nu+1}(x_2x_1+\omega_{12}\nu)}{\sqrt{\nu+x_1^2}((1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2)^{3/2}}\left[((1-\omega_{12}^2)(\nu+x_1^2)+(x_2-\omega_{12}x_1)^2)\right.\\
&\times\left(\omega_{12}\frac{\omega_{23}-\omega_{12}\omega_{13}}{1-\omega_{12}^2}-\omega_{13}\right)-\left((x_3-\omega_{13}x_1)-\frac{\omega_{23}-\omega_{12}\omega_{13}}{1-\omega_{12}^2}(x_2-\omega_{12}x_1)\right)(x_1-\omega_{12}x_2)\Big]\\
&+\psi\left(\frac{\sqrt{\nu+2}\left[(x_2-\omega_{12}x_1)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})(x_3-\omega_{13}x_1)\right]}{\sqrt{\left[(1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2\right]\left[(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2\right]}};\nu+2\right)\\
&\times\psi\left(v_{3\cdot1};\nu+1\right)\sqrt{\frac{(1-\omega_{13}^2)(\nu+2)}{(1-\omega_{12}^2)(1-\omega_{13}^2)-(\omega_{23}-\omega_{12}\omega_{13})^2}}\\
&\times\frac{\sqrt{\nu+1}(x_3x_1+\omega_{13}\nu)}{\sqrt{\nu+x_1^2}((1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2)^{3/2}}\left[((1-\omega_{13}^2)(\nu+x_1^2)+(x_3-\omega_{13}x_1)^2)\right.\\
&\times\left.\left(\omega_{13}\frac{\omega_{23}-\omega_{12}\omega_{13}}{1-\omega_{13}^2}-\omega_{12}\right)-\left((x_2-\omega_{12}x_1)-\frac{\omega_{23}-\omega_{12}\omega_{13}}{1-\omega_{13}^2}(x_3-\omega_{13}x)\right)(x_1-\omega_{13}x_3)\right].
\end{aligned}
$$

Combining the derivatives of the t cdf with equations (D.6)-(D.9) provides the full d-dimensional densities of the extremal-t process. Returning to the extremal skew-t case (i.e. when $\alpha \neq 0$ and $\tau \neq 0$), it is sufficient to consider the following changes. Firstly, rewrite

$$
T_j = \frac{\Psi_d\left\{\begin{pmatrix} u_j \\ \bar{\tau}_j \end{pmatrix};\begin{pmatrix} \bar{\Omega}_j^\circ & -\delta_j \\ -\delta_j^\top & 1 \end{pmatrix},\nu+1\right\}}{\Psi_1\left(\bar{\tau}_j;\nu+1\right)}, \quad j\in I,
$$

where $u_j = \left[\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left\{\left(\frac{x_i^\circ}{x_j^\circ}\right)^{1/\nu} - \omega_{i,j}\right\}, i \in I_j\right]^\top$, following Definition 4.1 . It can then be shown that

$$V_{1\cdots d} = -(\nu x_1)^{-(d+1)}\psi_{d-1}(u_1; \bar{\Omega}_1^\circ, \alpha_1^\circ, \tau_1^\circ, \kappa_1^\circ, \nu+1)\prod_{i=2}^{d}\sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}}\left(\frac{x_i^\circ}{x_1^\circ}\right)^{\frac{1}{\nu}-1}\frac{m_i^+}{m_1^+}$$

following Theorem 4.7. Note that equations (D.6)–(D.9) are still valid in this case, through the redefinition of $d \leftarrow d+1$ and $u_j \leftarrow (u_j, \bar{\tau}_j)^\top$. This in combination with the above derivatives of the t cdfs leads to the d-dimensional densities of the extremal-skew-t process.

D.3 Efficiency study: comparison between pairwise and triplewise extremal-t likelihood

We compare the efficiency of the maximum triplewise composite likelihood estimator with that based on the pairwise composite likelihood, discussed in Section 4 of the paper, when data are drawn from an extremal-t process. We generate 300 replicate samples of size $n = 20, 50$ and 70 from the extremal-t process with correlation function (4.10) in Section 4.2.2, with varying parameters, over 20 random spatial points on $\mathbb{S} = [0, 100]^2$. Table D.1 presents the resulting relative efficiencies $RE_\xi/RE_\lambda/RE_{(\lambda,\xi)}$ ($\times 100$), where $RE_\xi = \widehat{\text{var}}(\hat{\xi}_3)/\widehat{\text{var}}(\hat{\xi}_2)$, $RE_\lambda = \widehat{\text{var}}(\hat{\lambda}_3)/\widehat{\text{var}}(\hat{\lambda}_2)$ and $RE_{(\lambda,\xi)} = \widehat{\text{cov}}(\hat{\lambda}_3, \hat{\xi}_3)/\widehat{\text{cov}}(\hat{\lambda}_2, \hat{\xi}_2)$, where $(\hat{\lambda}_p, \hat{\xi}_p)$ are the p-wise maximum composite likelihood estimates ($p = 2, 3$), and $\widehat{\text{var}}$ and $\widehat{\text{cov}}$ denote sample variance and covariance over replicates. Perhaps unsurprisingly, the triplewise estimates are at worst just as efficient as the pairwise estimates ($RE \leq 100$) but are frequently much more efficient. However this is balanced computationally as there is a corresponding increase in the number of components in the triplewise composite likelihood function. For each ν, there is a general gain in efficiency when the smoothing parameter ξ increases for each fixed scale parameter λ. There is a similar gain when increasing λ for fixed ξ. These gains become progressively pronounced with increasing sample size n, and when there is stronger dependence present (i.e. smaller degrees of freedom ν). However, we note that there are a number of instances where the efficiency gain goes against this general trend, which indicates that there are some subtleties involved.

		$\nu = 1$			
$n = 20$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	89/94/89	84/97/93	83/69/79	81/82/84	78/64/72
28	76/100/98	59/100/69	73/86/73	74/66/75	34/75/26
42	81/100/100	51/96/89	51/80/88	43/63/79	33/51/72
$n = 50$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	85/81/84	87/78/86	76/67/78	66/56/72	52/47/62
28	64/100/81	81/79/82	73/72/78	72/66/74	34/68/24
42	71/100/97	33/61/59	17/42/40	17/34/37	2/18/7
$n = 70$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	80/87/83	81/76/80	74/65/77	62/57/70	47/42/60
28	51/100/68	82/82/84	72/72/77	71/66/73	54/53/62
42	56/93/89	28/52/48	13/40/14	12/28/27	8/23/26
		$\nu = 3$			
$n = 20$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	93/100/96	93/96/91	88/84/83	84/83/84	78/77/82
28	86/100/100	72/97/75	90/91/89	87/85/86	39/78/50
42	78/100/100	72/97/100	58/71/74	51/68/95	44/58/84
$n = 50$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	91/85/89	92/89/92	86/81/88	82/78/86	64/64/74
28	70/100/81	74/87/63	83/81/84	80/74/82	77/75/81
42	69/100/100	47/70/75	36/53/64	30/40/61	38/32/33
$n = 70$					
$\lambda\backslash\xi$	0.5	1	1.5	1.9	2
14	93/93/94	89/88/87	81/77/85	81/74/84	58/58/71
28	94/94/94	85/87/89	81/77/86	79/75/82	81/77/84
42	65/94/95	44/57/62	29/45/49	25/35/50	20/28/38

Table D.1: Efficiency of maximum triplewise likelihood estimators relative to maximum pairwise likelihood estimators for the Extremal-t process, based on 300 replicate simulations. Simulated datasets of size $n = 20, 50, 70$ are generated at 20 random sites in $\mathbb{S} = [0, 100]^2$, given power exponential dependence function parameters $\vartheta = (\lambda, \xi)$. Relative efficiencies are $RE_\xi/RE_\lambda/RE_{(\lambda,\xi)}$ ($\times 100$) where $RE_\xi = \widehat{\text{var}}(\hat{\xi}_3)/\widehat{\text{var}}(\hat{\xi}_2)$, $RE_\lambda = \widehat{\text{var}}(\hat{\lambda}_3)/\widehat{\text{var}}(\hat{\lambda}_2)$ and $RE_{(\lambda,\xi)} = \widehat{\text{cov}}(\hat{\lambda}_3, \hat{\xi}_3)/\widehat{\text{cov}}(\hat{\lambda}_2, \hat{\xi}_2)$, where $(\hat{\lambda}_p, \hat{\xi}_p)$ are the p-wise maximum composite likelihood estimates ($p = 2, 3$), and $\widehat{\text{var}}$ and $\widehat{\text{cov}}$ denote sample variance and covariance over replicates.

Appendix E

Chapter 5: Proofs

E.1 Preliminaries

The assumptions which will be used to establish the optimality properties of these kernel estimators are stated below. They are usually expressed for random variables with unbounded support, which in our case is the transformed variable $\boldsymbol{Y}$. This set of conditions do not form a minimal set, but they serve as a convenient starting point to state our results.

(A1) The d-variate density $f_{\boldsymbol{Y}}$ admits continuous, square integrable and ultimately monotone for all element-wise partial second derivatives.

(A2) The d-variate kernel is a positive, symmetric, square integrable p.d.f. such that $\int_{\mathbb{R}^d} \boldsymbol{y}\boldsymbol{y}^T K(\boldsymbol{y}) d\boldsymbol{y} = m_2(K)\mathbf{I}_d$ where $m_2(K)$ is finite and $\mathbf{I}_d$ is the $d \times d$ identity matrix.

(A3) The bandwidth matrix $\mathbf{H} = \mathbf{H}(n)$ forms a sequence of symmetric and positive definite matrices such that $n^{-1}|\mathbf{H}|^{-1/2}$ and every element of $\mathbf{H}$ approaches zero as $n \to \infty$.

The proof of Theorem 5.1 requires Lemma E.1 which establishes the minimal rate of MISE convergence of $\hat{f}_{\boldsymbol{Y}}$. This theorem has already been established, e.g. Wand (1992), but we include the details of a proof using an alternative notation for fourth order derivatives of a multivariate function as four-fold Kronecker product, see Holmquist (1996).

Lemma E.1. *Suppose that the conditions (A1–A3) hold. The minimal MISE of the transformed kernel density estimator* $\hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H})$ *is*

$$\begin{aligned} \inf_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) &= \left[\tfrac{1}{4} m_2^2(K) (\operatorname{vec}^T \mathbf{H}^*)^{\otimes 2} \boldsymbol{\psi}_{\boldsymbol{Y},4} + n^{-1} |\mathbf{H}^*|^{-1/2} R(K)\right] \{1 + o(1)\} \\ &= O(n^{-4/(d+4)}) \end{aligned}$$

where $\boldsymbol{\psi}_{\boldsymbol{Y},4} = \int_{\mathbb{R}^d} \mathsf{D}^{\otimes 4} f_{\boldsymbol{Y}}(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$.

Proof of Lemma E.1. The expected value of $\hat{f}_{\boldsymbol{Y}}$ is

$$\mathbb{E}\, \hat{f}_{\boldsymbol{Y}}(\boldsymbol{y}; \mathbf{H}) = \mathbb{E}\, K_{\mathbf{H}}(\boldsymbol{y} - \boldsymbol{Y}) = \int_{\mathbb{R}^d} K_{\mathbf{H}}(\boldsymbol{y} - \boldsymbol{w}) f(\boldsymbol{y})\, d\boldsymbol{w} = K_{\mathbf{H}} * f_{\boldsymbol{Y}}(x)$$

where $*$ is the convolution operator between two functions. Asymptotically, using a Taylor series expansion and the usual rearrangement $\boldsymbol{w}^T\mathbf{H}^{1/2}\mathsf{D}^2F(\boldsymbol{x})\mathbf{H}^{1/2} = \operatorname{tr}(\boldsymbol{w}\boldsymbol{w}^T\mathbf{H}\mathsf{D}^2F(\boldsymbol{x}))$,

$$\begin{aligned}
\mathbb{E}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H}) &= \int_{\mathbb{R}^d} |\mathbf{H}|^{-1/2}K(\mathbf{H}^{-1/2}(\boldsymbol{y}-\boldsymbol{w}))f_{\boldsymbol{Y}}(\boldsymbol{y})d\boldsymbol{w} \\
&= \int_{\mathbb{R}^d} K(\boldsymbol{w})f(\boldsymbol{y}-\mathbf{H}^{1/2}\boldsymbol{w})d\boldsymbol{w} \\
&= \int_{\mathbb{R}^d} K(\boldsymbol{w})[f(\boldsymbol{y}) - \boldsymbol{w}^T\mathbf{H}^{1/2}\mathsf{D}f(\boldsymbol{x}) + \tfrac{1}{2}\boldsymbol{w}^T\mathbf{H}^{1/2}\mathsf{D}^2f(\boldsymbol{y})\mathbf{H}^{1/2}\boldsymbol{w}]\{1+o(1)\}\,d\boldsymbol{w} \\
&= [f(\boldsymbol{y}) + \tfrac{1}{2}\int K(\boldsymbol{w})\operatorname{tr}(\boldsymbol{w}\boldsymbol{w}^T\mathbf{H}\mathsf{D}^2f(\boldsymbol{y}))\,d\boldsymbol{w}]\{1+o(1)\} \\
&= [f(\boldsymbol{y}) + \tfrac{1}{2}m_2(K)\operatorname{tr}(\mathbf{H}\mathsf{D}^2f(\boldsymbol{y}))]\{1+o(1)\}.
\end{aligned}$$

This allows us to write the bias of $\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H})$ as

$$\mathbb{E}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H}) - f(\boldsymbol{y}) = \tfrac{1}{2}m_2(K)\operatorname{tr}(\mathbf{H}\mathsf{D}^2f(\boldsymbol{y}))\{1+o(1)\}.$$

For the variance, we have $\operatorname{Var}\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H}) = n^{-1}\,\mathbb{E}[K_{\mathbf{H}}(\boldsymbol{y}-\boldsymbol{Y})^2] - n^{-1}[\mathbb{E}\,K_{\mathbf{H}}(\boldsymbol{y}-\boldsymbol{Y})]^2$. The second term is the same as above, so it leaves us to evaluate

$$\begin{aligned}
\mathbb{E}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H})^2 = \mathbb{E}[K_{\mathbf{H}}(\boldsymbol{y}-\boldsymbol{Y})^2] &= \int_{\mathbb{R}^d} K_{\mathbf{H}}(\boldsymbol{y}-\boldsymbol{w})^2 f_{\boldsymbol{Y}}(\boldsymbol{w})\,d\boldsymbol{w} \\
&= \int_{\mathbb{R}^d} |\mathbf{H}|^{-1}K(\mathbf{H}^{-1/2}(\boldsymbol{y}-\boldsymbol{w}))^2 f_{\boldsymbol{Y}}(\boldsymbol{w})\,d\boldsymbol{w} \\
&= \int_{\mathbb{R}^d} |\mathbf{H}|^{-1/2}K(\boldsymbol{w})^2 f_{\boldsymbol{Y}}(\boldsymbol{y}-\mathbf{H}^{-1/2}\boldsymbol{w})d\boldsymbol{w} \\
&= |\mathbf{H}|^{-1/2}f_{\boldsymbol{Y}}(\boldsymbol{y})\int_{\mathbb{R}^d} K(\boldsymbol{w})^2 d\boldsymbol{w}\{1+o(1)\} \\
&= |\mathbf{H}|^{-1/2}f_{\boldsymbol{Y}}(\boldsymbol{y})R(K)\{1+o(1)\}.
\end{aligned}$$

Thus the variance term is

$$\operatorname{Var}\hat{f}(\boldsymbol{y};\mathbf{H}) = n^{-1}\{|\mathbf{H}|^{-1/2}f_{\boldsymbol{Y}}(\boldsymbol{y})R(K) - [f_{\boldsymbol{Y}}(\boldsymbol{y}) + \tfrac{1}{2}m_2(K)\operatorname{tr}(\mathbf{H}\mathsf{D}^2f_{\boldsymbol{Y}}(\boldsymbol{y}))]^2\}\{1+o(1)\}.$$

Since $\mathbf{H}\to 0$ then $|\mathbf{H}|^{-1/2}$ dominates both the constant term $f_{\boldsymbol{Y}}(\boldsymbol{y})$ and the $\operatorname{tr}(\mathbf{H})$ term so we can write

$$\operatorname{Var}\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H}) = n^{-1}|\mathbf{H}|^{-1/2}f_{\boldsymbol{Y}}(\boldsymbol{y})R(K)\{1+o(1)\}.$$

The integrated square bias (ISB) is

$$\begin{aligned}
\operatorname{ISB}\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}) = \int_{\mathbb{R}^d}\operatorname{Bias}^2\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H})\mathrm{d}\boldsymbol{y} &= \int_{\mathbb{R}^d}\tfrac{1}{4}m_2^2(K)\operatorname{tr}^2(\mathbf{H}\mathsf{D}^2f(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}\{1+o(1)\} \\
&= \tfrac{1}{4}m_2^2(K)\int_{\mathbb{R}^d}\operatorname{tr}^2(\mathbf{H}\mathsf{D}^2f(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}\{1+o(1)\} \\
&= \tfrac{1}{4}m_2^2(K)(\operatorname{vec}^T\mathbf{H}\otimes\operatorname{vec}^T\mathbf{H})\boldsymbol{\psi}_{\boldsymbol{Y},4}\{1+o(1)\},
\end{aligned}$$

and similarly the integrated variance (IV) is

$$\begin{aligned}\mathrm{IV}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}) &= \int_{\mathbb{R}^d} n^{-1}|\mathbf{H}|^{-1/2} f_{\boldsymbol{Y}}(\boldsymbol{y}) R(K)\mathrm{d}\boldsymbol{y}\{1+o(1)\}\\ &= n^{-1}|\mathbf{H}|^{-1/2}R(K)\int_{\mathbb{R}^d} f_{\boldsymbol{Y}}(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\{1+o(1)\}\\ &= n^{-1}|\mathbf{H}|^{-1/2}R(K)\{1+o(1)\}.\end{aligned}$$

using the integrability assumptions in conditions (A1) and (A2). Hence we obtain

$$\begin{aligned}\mathrm{MISE}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}) &= \mathrm{ISB}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}) + \mathrm{IV}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H})\\ &= [\tfrac{1}{4}m_2^2(K)(\mathrm{vec}^T\,\mathbf{H}\otimes\mathrm{vec}^T\,\mathbf{H})\boldsymbol{\psi}_{\boldsymbol{Y},4} + n^{-1}|\mathbf{H}|^{-1/2}R(K)]\{1+o(1)\}.\end{aligned}$$

The optimiser of this MISE requires that the order of the squared bias and the variance are the same, so then $\mathbf{H}^* = O(n^{-2/(d+4)})$, and so

$$\begin{aligned}\inf_{\mathbf{H}\in\mathcal{F}}\mathrm{MISE}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}) &= \mathrm{MISE}\,\hat{f}_{\boldsymbol{Y}}(\cdot;\mathbf{H}^*)\\ &= [\tfrac{1}{4}m_2^2(K)(\mathrm{vec}^T\,\mathbf{H}^*\otimes\mathrm{vec}^T\,\mathbf{H}^*)\boldsymbol{\psi}_{\boldsymbol{Y},4} + n^{-1}|\mathbf{H}^*|^{-1/2}R(K)]\{1+o(1)\}\\ &= O(n^{-4/(d+4)}).\end{aligned}$$

□

E.2 Proof of Theorem 5.1

Firstly using the definition of $\hat{f}_{\boldsymbol{X}^{[u]}}$ given in (5.1), its expected value is $\mathbb{E}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) = |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|\,\mathbb{E}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H})$ and its associated bias is

$$\mathrm{Bias}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) = \mathbb{E}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) - f_{\boldsymbol{X}^{[u]}}(\boldsymbol{x}) = |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|\,\mathrm{Bias}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H}).$$

Similarly we have $\mathbb{E}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H})^2 = |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|^2\,\mathbb{E}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H})^2$, leading to

$$\mathrm{Var}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) = \mathbb{E}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H})^2 - [\,\mathbb{E}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H})]^2 = |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|^2\,\mathrm{Var}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H}).$$

The integrated square bias (ISB) is

$$\begin{aligned}\mathrm{ISB}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) &= \int_{(\boldsymbol{u},\infty)} |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|^2\,\mathrm{Bias}^2\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H})\mathrm{d}\boldsymbol{x}\\ &= \int_{\mathbb{R}^d} |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{t}^{-1}(\boldsymbol{y}))|^2\,\mathrm{Bias}^2\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{t}^{-1}(\boldsymbol{y}));\mathbf{H})\mathrm{d}\boldsymbol{y}.\end{aligned}$$

given the change of variable $\boldsymbol{y} = \boldsymbol{t}(\boldsymbol{x})$. Similarly the integrated variance (IV) is

$$\mathrm{IV}\,\hat{f}_{\boldsymbol{X}^{[u]}}(\boldsymbol{x};\mathbf{H}) = \int_{(\boldsymbol{u},\infty)} |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{x})|^2\,\mathrm{Var}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{t}(\boldsymbol{x});\mathbf{H})\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{t}^{-1}(\boldsymbol{y}))|^2\,\mathrm{Var}\,\hat{f}_{\boldsymbol{Y}}(\boldsymbol{y};\mathbf{H})\mathrm{d}\boldsymbol{x}.$$

As a consequence, the MISE of $\hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ can be written as

$$\operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H}) = \int_{\mathbb{R}^d} |\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{t}^{-1}(\boldsymbol{y}))| \operatorname{MSE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) \mathrm{d}\boldsymbol{y}.$$

Since the $|\mathbf{J}_{\boldsymbol{t}}(\boldsymbol{t}^{-1}(\boldsymbol{y}))|$ term does not involve the sample size n or the bandwidth matrix $\mathbf{H}$, it has no influence on the rate of convergence, thus the MISE of $\hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}$ and $\hat{f}_{\boldsymbol{Y}}$ are asymptotically equivalent, which we summarize by

$$\operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H}) = \{\operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H})\}\{1 + o(1)\}.$$

Since $\mathbf{H}^*$ and $\mathbf{H}^{*[\boldsymbol{u}]} = \operatorname{argmin}_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H})$ are of the same asymptotic $n^{-2/(d+4)}$ order, and using Lemma E.1, we then have

$$\begin{aligned}
&\inf_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H}) - \Big\{ \inf_{\mathbf{H} \in \mathcal{F}} \operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}) \Big\} = \operatorname{MISE} \hat{f}_{\boldsymbol{X}^{[\boldsymbol{u}]}}(\cdot; \mathbf{H}^{*[\boldsymbol{u}]}) - \operatorname{MISE} \hat{f}_{\boldsymbol{Y}}(\cdot; \mathbf{H}^*) \\
&\quad = \{ \tfrac{1}{4} m_2^2(K) [(\operatorname{vec}^T \mathbf{H}^{*[\boldsymbol{u}]})^{\otimes 2} \boldsymbol{\psi}_{\boldsymbol{X}^{[u]},4} - (\operatorname{vec}^T \mathbf{H}^*)^{\otimes 2} \boldsymbol{\psi}_{\boldsymbol{Y},4}] \\
&\quad + n^{-1} (|\mathbf{H}^{*[\boldsymbol{u}]}|^{-1/2} - |\mathbf{H}^*|^{-1/2}) R(K) \} \{1 + o(1)\} = O(n^{-4/(d+4)}).
\end{aligned}$$

List of Figures

List of Tables

Bibliography

Abramowitz, M. and Stegun, I. A. (1964), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of *National Bureau of Standards Applied Mathematics Series*, For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Apputhurai, P. and Stephenson, A. (2011), "Accounting for uncertainty in extremal dependence modeling using bayesian model averaging techniques," *Journal of Statistical Planning and Inference*, 141, 1800–1807.

Arellano-Valle, R. B. and Azzalini, A. (2006), "On the unification of families of skew-normal distributions," *Scandinavian Journal of Statistics*, 561–574.

Arellano-Valle, R. B. and Genton, M. G. (2010), "Multivariate extended skew-t distributions and related families," *Metron*, 68, 201–234.

Azzalini, A. (1985), "A class of distributions which includes the normal ones," *Scandinavian journal of statistics*, 171–178.

— (2005), "The skew-normal distribution and related multivariate families," *Scand. J. Statist.*, 32, 159–200, with discussion by Marc G. Genton and a rejoinder by the author.

— (2013), *The skew-normal and related families*, vol. 3, Cambridge University Press.

Azzalini, A. and Dalla Valle, A. (1996), "The multivariate skew-normal distribution," *Biometrika*, 83, 715–726.

Balkema, A. A. and de Haan, L. (1974), "Residual life time at great age," *Ann. Probability*, 2, 792–804.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006), *Statistics of extremes: theory and applications*, John Wiley & Sons.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004), *Statistics of extremes*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, theory and applications, With contributions from Daniel De Waal and Chris Ferro.

Beranger, B., Marcon, G., and Padoan, S. (2015), *ExtremalDep: Extremal Dependence Modeling*, r package version 0.1-2/r76.

Beranger, B. and Padoan, S. A. (2015), "Extreme Dependence Models," *ArXiv e-prints*.

Beranger, B., Padoan, S. A., and Sisson, S. A. (2015), "Models for extremal dependence derived from skew-symmetric families," *ArXiv e-prints*.

Billingsley, P. (1968), *Convergence of probability measures*, John Wiley & Sons, Inc., New York-London-Sydney.

— (1971), *Weak convergence of measures: Applications in probability*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

Blanchet, J. and Davison, A. C. (2011), "Spatial modeling of extreme snow depth," *Ann. Appl. Stat.*, 5, 1699–1725.

Boldi, M.-O. and Davison, A. C. (2007), "A mixture model for multivariate extremes," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69, 217–229.

Bortot, P. (2010), "Tail dependence in bivariate skew-normal and skew-t distributions," .

Bowman, A. W., Hall, P., and Titterington, D. M. (1984), "Cross-validation in nonparametric estimation of probabilities and probability densities," *Biometrika*, 71, 341–351.

Brown, B. M. and Resnick, S. I. (1977), "Extreme values of independent stochastic processes," *Journal of Applied Probability*, 732–739.

Buishand, T. (1984), "Bivariate extreme-value data and the station-year method," *Journal of Hydrology*, 69, 77 – 95.

Buishand, T. A., de Haan, L., and Zhou, C. (2008), "On spatial extremes: with application to a rainfall problem," *Ann. Appl. Stat.*, 2, 624–642.

Capéraà, P. and Fougères, A.-L. (2000), "Estimation of a bivariate extreme value distribution," *Extremes*, 3, 311–329 (2001).

Capitanio, A., Azzalini, A., and Stanghellini, E. (2003), "Graphical models for skew-normal variates," *Scand. J. Statist.*, 30, 129–144.

Chan, G. and Wood, A. T. (1997), "Algorithm AS 312: An Algorithm for simulating stationary Gaussian random fields," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46, 171–181.

Chang, S.-M. and Genton, M. G. (2007), "Extreme value distributions for the skew-symmetric family of distributions," *Communications in Statistics—Theory and Methods*, 36, 1705–1717.

Chen, S. X. (1999), "Beta kernel estimators for density functions," *Comput. Statist. Data Anal.*, 31, 131–145.

Cheng, R. C. H. and Amin, N. A. K. (1983), "Estimating parameters in continuous univariate distributions with a shifted origin," *J. Roy. Statist. Soc. Ser. B*, 45, 394–403.

Chilès, J.-P. and Delfiner, P. (1999), *Geostatistics*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, modeling spatial uncertainty, A Wiley-Interscience Publication.

Christopeit, N. (1994), "Estimating parameters of an extreme value distribution by the method of moments," *J. Statist. Plann. Inference*, 41, 173–186.

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer-Verlag London, Ltd., London.

Coles, S. G. and Tawn, J. A. (1991), "Modelling Extreme Multivariate Events," *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, pp. 377–392.

— (1994), "Statistical Methods for Multivariate Extremes: An Application to Structural Design," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, pp. 1–48.

Cooley, D., Davis, R. A., and Naveau, P. (2010), "The pairwise beta distribution: a flexible parametric multivariate model for extremes," *J. Multivariate Anal.*, 101, 2103–2117.

Cooley, D., Naveau, P., and Poncet, P. (2006), "Variograms for spatial max-stable random fields," in *Dependence in probability and statistics*, New York: Springer, vol. 187 of *Lecture Notes in Statist.*, pp. 373–390.

Daley, D. J. and Vere-Jones, D. (1988), *An introduction to the theory of point processes*, Springer Series in Statistics, Springer-Verlag, New York.

Davison, A. C. (2003), *Statistical Models*, Cambridge.

Davison, A. C. and Gholamrezaee, M. M. (2012), "Geostatistics of extremes," *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 468, 581–608.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), "Statistical Modeling of Spatial Extremes," *Statistical Science*, 27, 161–186.

de Carvalho, M., Oumow, B., Segers, J., and Warchoł, M. (2013), "A Euclidean likelihood estimator for bivariate tail dependence," *Comm. Statist. Theory Methods*, 42, 1176–1192.

de Haan, L. (1984), "A spectral representation for max-stable processes," *Ann. Probab.*, 12, 1194–1204.

de Haan, L. and Ferreira, A. (2006), *Extreme value theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York, an introduction.

de Haan, L. and Pereira, T. T. (2006), "Spatial extremes: models for the stationary case," *Ann. Statist.*, 34, 146–168.

de Haan, L. and Resnick, S. I. (1977), "Limit theory for multivariate sample extremes," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40, 317–337.

de Haan, L. and Zhou, C. (2008), "On extreme value analysis of a spatial process," *REVSTAT*, 6, 71–81.

De Haan, L. and Zhou, C. (2011), "Extreme residual dependence for random vectors and processes," *Advances in Applied Probability*, 43, 217–242.

Demarta, S. and McNeil, A. J. (2005), "The t Copula and Related Copulas," *International Statistical Review*, 73, 111–129.

Devroye, L. and Györfi, L. (1985), *Nonparametric Density Estimation: the L_1 View*, New York: John Wiley and Sons.

Dieker, A. and Mikosch, T. (2015), "Exact simulation of Brown-Resnick random fields at a finite number of locations," *Extremes*, 18, 301–314.

Dombry, C. and Eyi-Minko, F. (2013), "Regular conditional distributions of continuous max-infinitely divisible random fields," *Electron. J. Probab.*, 18, no. 7, 21.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013), "Conditional simulation of max-stable processes," *Biometrika*, 100, 111–124.

Drees, H. and Huang, X. (1998), "Best attainable rates of convergence for estimators of the stable tail dependence function," *J. Multivariate Anal.*, 64, 25–47.

Duong, T. and Hazelton, M. L. (2003), "Plug-in bandwidth matrices for bivariate kernel density estimation," *J. Nonparametr. Stat.*, 15, 17–30.

— (2005), "Cross-validation bandwidth matrices for multivariate kernel density estimation," *Scand. J. Statist.*, 32, 485–506.

Dutt, J. E. (1973), "A representation of multivariate normal probability integrals by integral transforms," *Biometrika*, 60, 637–645.

Einmahl, J. H. J., de Haan, L., and Piterbarg, V. I. (2001), "Nonparametric estimation of the spectral measure of an extreme value distribution," *Ann. Statist.*, 29, 1401–1423.

Einmahl, J. H. J., Krajina, A., and Segers, J. (2008), "A method of moments estimator of tail dependence," *Bernoulli*, 14, 1003–1026.

— (2012), "An M-estimator for tail dependence in arbitrary dimensions," *Ann. Statist.*, 40, 1764–1793.

Einmahl, J. H. J. and Segers, J. (2009), "Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution," *Ann. Statist.*, 37, 2953–2989.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling extremal events*, vol. 33 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin, for insurance and finance.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2012), "Estimation of Huesler-Reiss distributions and Brown-Resnick processes," *ArXiv e-prints*.

Erhardt, R. J. and Smith, R. L. (2012), "Approximate Bayesian computing for spatial extremes," *Comput. Statist. Data Anal.*, 56, 1468–1481.

Falk, M. and Guillou, A. (2008), "Peaks-over-threshold stability of multivariate generalized Pareto distributions," *J. Multivariate Anal.*, 99, 715–734.

Falk, M., Hüsler, J., and Reiss, R.-D. (2011), *Laws of small numbers: extremes and rare events*, Birkhäuser/Springer Basel AG, Basel, extended ed.

Feller, W. (1968), "An Introduction to Probability Theory, Vol. 1," .

Ferreira, A. and de Haan, L. (2014), "The generalized Pareto process; with a view towards application and simulation," *Bernoulli*, 20, 1717–1737.

Fisher, R. A. and Tippett, L. H. C. (1928), "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013), *Evaluation of Climate Models*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, book section 9, pp. 741–866.

Gasser, T. and Müller, H.-G. (1979), "Kernel estimation of regression functions," in *Smoothing Techniques for Curve Estimation*, eds. Gasser, T. and Rosenblatt, M., Berlin: Springer, pp. 23–68.

Geenens, G. (2014), "Probit Transformation for Kernel Density Estimation on the Unit Interval," *J. Amer. Statist. Assoc.*, 109, 346–358.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010), *Handbook of spatial statistics*, CRC press.

Genton, M. (2004), *Skew-elliptical distributions and their applications*, Chapman & Hall/CRC, Boca Raton, FL, a journey beyond normality, Edited by Marc G. Genton.

Genton, M. G., Ma, Y., and Sang, H. (2011), "On the likelihood function of Gaussian max-stable processes," *Biometrika*, 98, 481–488.

Genton, M. G., Padoan, S. A., and Sang, H. (2015), "Multivariate max-stable spatial processes," *Biometrika*, 102, 215–230.

Geweke, J. (1992), *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, vol. Bayesian Statistics, University Press.

Giné, E., Hahn, M. G., and Vatan, P. (1990), "Max-infinitely divisible and max-stable sample continuous processes," *Probab. Theory Related Fields*, 87, 139–165.

Gnedenko, B. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire," *Ann. of Math. (2)*, 44, 423–453.

Gudendorf, G. and Segers, J. (2011), "Nonparametric estimation of an extreme-value copula in arbitrary dimensions," *Journal of Multivariate Analysis*, 102.

— (2012), "Nonparametric estimation of multivariate extreme-value copula," *Journal of Statistical Planning and Inference*, 142.

Gumbel, E. J. (1960), "Distributions des valeurs extremes en plusieurs dimensions," *Publ. Inst. Statist. Univ. Paris*, 9.

Hall, P., Marron, J. S., and Park, B. U. (1992), "Smoothed cross-validation," *Probab. Theory Related Fields*, 92, 1–20.

Hall, P. and Tajvidi, N. (2000), "Distribution and dependence-function estimation for bivariate extreme-value distributions," *Bernoulli*, 6, 835–844.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, pp. 97–109.

Heffernan, J. E. and Tawn, J. A. (2004), "A conditional approach for multivariate extreme values," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66, 497–546, with discussions and reply by the authors.

Heidelberger, P. and Welch, P. D. (1981), "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," *Commun. ACM*, 24, 233–245.

Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution," *Ann. Statist.*, 3, 1163–1174.

Holmquist, B. (1996), "The d-variate vector Hermite polynomial of order k," *Linear Algebra and Its Applications*, 237/238, 155–190.

Hosking, J. R. M. (1985), "Algorithm AS 215: Maximum-Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution," *J. Roy. Stat. Soc. Ser. C*, 34, pp. 301–310.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985), "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments," *Technometrics*, 27, 251–261.

Huang, X. (1992), "Statistics of bivariate extreme values," Dissertation, thesis, zugl.: Rotterdam, Univ., Diss., 1991.

Huser, R. and Davison, A. C. (2013), "Composite likelihood estimation for the Brown-Resnick process," *Biometrika*, 100, 511–518.

Huser, R. and Genton, M. (2015), "Non-stationary dependence structures for spatial extremes," *arXiv:1411.3174v1*.

Hüsler, J. and Reiss, R.-D. (1989), "Maxima of normal random vectors: between independence and complete dependence," *Statist. Probab. Lett.*, 7, 283–286.

Jamalizadeh, A., Mehrali, Y., and Balakrishnan, N. (2009), "Recurrence relations for bivariate t and extended skew-t distributions and an application to order statistics from bivariate t," *Computational Statistics & Data Analysis*, 53, 4018–4027.

Jenkinson, A. F. (1955), "The frequency distribution of the annual maximum (or minimum) values of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, 81, 158–171.

Joe, H. (1990), "Families of min-stable multivariate exponential and multivariate extreme value distributions," *Statist. Probab. Lett.*, 9, 75–81.

— (1997), *Multivariate models and dependence concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Johansen, S. S. (2004), "Bivariate frequency analysis of flood characteristics in Glomma and Gudbrandsdalslagen," Ph.D. thesis, University of Oslo Department of Geosciences Section of Geohazards and Hydrology.

Jones, D. A., Wang, W., and Fawcett, R. (2009), "High-quality spatial climate data-sets for Australia," *Australian Meteorological and Oceanographic Journal*, 58, 233–1026.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009), "Stationary max-stable fields associated to negative definite functions," *Ann. Probab.*, 37, 2042–2065.

Kotz, S. and Nadarajah, S. (2000), *Extreme value distributions*, London: Imperial College Press, theory and applications.

Ledford, A. W. and Tawn, J. A. (1996), "Statistics for near independence in multivariate extreme values," *Biometrika*, 83, 169–187.

— (1997), "Modelling dependence within joint tail regions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 475–499.

Li, H. (2009), "Orthant tail dependence of multivariate extreme value distributions," *Journal of Multivariate Analysis*, 100, 243–256.

Lindgren, G. (2012), *Stationary Stochastic Processes: Theory and Applications*, CRC Press.

Lindsay, B. G. (1988), "Composite likelihood methods," in *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, Amer. Math. Soc., Providence, RI, vol. 80 of *Contemp. Math.*, pp. 221–239.

Lye, L., Hapuarachchi, K., and Ryan, S. (1993), "Bayes estimation of the extreme-value reliability function," *IEEE Trans. Reliab.*, 42, 641–644.

Lysenko, N., Roy, P., and Waeber, R. (2009), "Multivariate extremes of generalized skew-normal distributions," *Statist. Probab. Lett.*, 79, 525–533.

Macleod, A. J. (1989), "Remark AS R76: A Remark on Algorithm AS 215: Maximum-Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution," *J. Roy. Statist. Soc. Ser. C*, 38, 198–199.

Mandelbrot, B. B. and Van Ness, J. W. (1968), "Fractional Brownian motions, fractional noises and applications," *SIAM review*, 10, 422–437.

Marcon, G., Padoan, S. A., Naveau, P., and Muliere, P. (2014), "Multivariate Nonparametric Estimation of the Pickands Dependence Function using Bernstein Polynomials," *Under review.*

Maritz, J. S. and Munro, A. H. (1967), "On the use of the generalised extreme-value distribution in estimating extreme percentiles," *Biometrics*, 23, 79–103.

Markovich, N. (2007), *Nonparametric Analysis of Univariate Heavy-tailed Data: Research and Practice*, Chichester: John Wiley & Sons.

Marshall, J. C. and Hazelton, M. L. (2010), "Boundary kernels for adaptive density estimators on regions with irregular boundaries," *J. Multivariate Anal.*, 101, 949–963.

Minozzo, M. and Ferracuti, L. (2012), "On the existence of some skew-normal stationary processes," *Chilean Journal of Statistics*, 3, 157–170.

Misés, R. D. (1936), "Les lois de probabilité pour les fonctions statistiques," *Ann. Inst. H. Poincaré*, 6, 185–212.

Molchanov, I. and Stucki, K. (2013), "Stationarity of multivariate particle systems," *Stochastic Processes and their Applications*, 123, 2272 – 2285.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009), "Modelling pairwise dependence of maxima in space," *Biometrika*, 96, 1–17.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009), "Extreme value properties of multivariate t copulas," *Extremes*, 12, 129–148.

Nolan, J. (2003), *Stable distributions: models for heavy-tailed data*, Birkhauser.

Oesting, M., Kabluchko, Z., and Schlather, M. (2012), "Simulation of Brown-Resnick processes," *Extremes*, 15, 89–107.

Oesting, M. and Schlather, M. (2014), "Conditional sampling for max-stable processes with a mixed moving maxima representation," *Extremes*, 17, 157–192.

Oesting, M., Schlather, M., and Zhou, C. (2013), "On the Normalized Spectral Representation of Max-Stable Processes on a Compact Set," *ArXiv e-prints*.

Opitz, T. (2013), "Extremal t processes: Elliptical domain of attraction and a spectral representation," *Journal of Multivariate Analysis*, 122, 409 – 413.

Padoan, S. A. (2011), "Multivariate extreme models based on underlying skew- and skew-normal distributions," *Journal of Multivariate Analysis*, 102, 977 – 991.

— (2013a), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd., Chichester, chap. Max-Stable Processes, no. 4.

— (2013b), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd., Chichester, chap. Extreme Value Analysis, no. 2.

— (2013c), "Extreme Dependence Models Based on Event Magnitude," *Journal of Multivariate Analysis*, 122, 1–19.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010), "Likelihood-Based Inference for Max-Stable Processes," *Journal of the American Statistical Association*, 105, 263–277.

Penrose, M. D. (1992), "Semi-min-stable processes," *Ann. Probab.*, 20, 1450–1463.

Perkins, S. E., Moise, A., Whetton, P., and Katzfey, J. (2014), "Regional changes of climate extremes over Australia Ð a comparison of regional dynamical downscaling and global climate model simulations," *International Journal of Climatology*, 34, 3456–3478.

Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J. (2007), "Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions," *J. Climate*, 20, 4356–4376.

Perkins, S. E., Pitman, A. J., and Sisson, S. A. (2013), "Systematic differences in future 20 year temperature extremes in AR4 model projections over Australia as a function of model skill," *Int. J. Climatol.*, 33, 1153–1167.

Pickands, III, J. (1975), "Statistical inference using extreme order statistics," *Ann. Statist.*, 3, 119–131.

— (1981), "Multivariate extreme value distributions," in *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, vol. 49, pp. 859–878, 894–902, with a discussion.

Prescott, P. and Walden, A. T. (1980), "Maximum likelihood estimation of the parameters of the generalized extreme-value distribution," *Biometrika*, 67, 723–724.

Resnick, S. I. (1987), *Extreme values, regular variation, and point processes*, vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*, New York: Springer-Verlag.

— (2007), *Extreme values, regular variation, and point processes*, Springer.

Ribatet, M. (2013), "Spatial extremes: Max-stable processes at work," *Journal de la Société Francaise de Statistique*, 154, 156–177.

Ribatet, M., Cooley, D., and Davison, A. C. (2012), "Bayesian inference from composite likelihoods, with an application to spatial extremes," *Statist. Sinica*, 22, 813–845.

Ribatet, M. and Sedki, M. (2013), "Extreme value copulas and max-stable processes," *J. SFdS*, 154, 138–150.

Ribatet, M., Singleton, R., and team, R. C. (2013), *SpatialExtremes: Modelling Spatial Extremes*, r package version 2.0-0.

Rootzén, H. and Tajvidi, N. (2006), "Multivariate generalized Pareto distributions," *Bernoulli*, 12, 917–930.

Rudemo, M. (1982), "Empirical choice of histograms and kernel density estimators," *Scand. J. Statist.*, 9, 65–78.

Ruppert, D. and Cline, D. B. H. (1994), "Transformation kernel density estimation – bias reduction by empirical transformations," *Ann. Statist.*, 22, 185–210.

Ruppert, D. and Wand, M. P. (1992), "Correcting for kurtosis in density estimation," *Austral. J. Statist.*, 34, 19–29.

Sabourin, A. and Naveau, P. (2014), "Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization," *Comput. Statist. Data Anal.*, 71, 542–567.

Sabourin, A., Naveau, P., and Fougères, A.-L. (2013), "Bayesian model averaging for multivariate extremes," *Extremes*, 16, 325–350.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, D. Reidel Publishing Company.

Samorodnitsky, G. and Taqqu, M. S. (1994), *Stable non-Gaussian random processes*, Stochastic Modeling, Chapman & Hall, New York, stochastic models with infinite variance.

Sang, H. and Gelfand, A. E. (2010), "Continuous spatial process models for spatial extreme values," *J. Agric. Biol. Environ. Stat.*, 15, 49–65.

Sang, H. and Genton, M. G. (2014), "Tapered composite likelihood for spatial max-stable models," *Spat. Stat.*, 8, 86–103.

Schilling, R. L. (2005), *Measures, integrals and martingales*, Cambridge University Press, New York.

Schlather, M. (2002), "Models for stationary max-stable random fields," *Extremes*, 5, 33–44.

Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Menck, P. J., Gross, S., Ober, U., Burmeister, K., Manitz, J., Ribeiro, P., Singleton, R., Pfaff, B., and R Core Team (2015), *RandomFields: Simulation and Analysis of Random Fields*, r package version 3.0.62.

Schlather, M. and Tawn, J. A. (2003), "A dependence measure for multivariate and spatial extreme values: properties and inference," *Biometrika*, 90, 139–156.

Scott, D. W. (2015), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Hoboken, NJ: John Wiley & Sons, 2nd ed.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X. (2012), *Changes in Climate Extremes and their Impacts on the Natural Physical Environment*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, book section 3, pp. 190–230.

Sheather, S. J. and Jones, M. C. (1991), "A reliable data-based bandwidth selection method for kernel density estimation," *J. Roy. Statist. Soc. Ser. B*, 53, 683–690.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D. (2013a), "Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate," *Journal of Geophysical Research: Atmospheres*, 118, 1716–1733.

Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., and Bronaugh, D. (2013b), "Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections," *Journal of Geophysical Research: Atmospheres*, 118, 2473–2493.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, New York: Springer-Verlag.

Smith, E. L. and Stephenson, A. G. (2009), "An extended Gaussian max-stable process model for spatial extremes," *J. Statist. Plann. Inference*, 139, 1266–1275.

Smith, R. L. (1985), "Maximum likelihood estimation in a class of nonregular cases," *Biometrika*, 72, 67–90.

— (1990), "Max-stable processes and spatial extremes," Tech. rep., University of Surrey.

Smith, R. L., Tawn, J. A., and Yuen, H. K. (1990), "Statistics of Multivariate Extremes," *Int. Statist. Rev.*, 58, 47–58.

Stoev, S. A. and Taqqu, M. S. (2005), "Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes," *Extremes*, 8, 237–266 (2006).

Tawn, J. A. (1990), "Modelling Multivariate Extreme Value Distributions," *Biometrika*, 77, pp. 245–253.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012), "An Overview of CMIP5 and the Experiment Design," *Bulletin of the American Meteorological Society*, 93, 485–498.

Team, R. D. C. (2013), *R: a Language and Environment for Statistical Computing*.

Varin, C., Reid, N., and Firth, D. (2011), "An overview of composite likelihood methods," *Statist. Sinica*, 21, 5–42.

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., Halenka, T., Karacostas, T., Katragkou, E., Keuler, K., Kotlarski, S., Mayer, S., van Meijgaard, E., Nikulin, G., Patarcić, M., Scinocca, J., Sobolowski, S., Suklitsch, M., Teichmann, C., Warrach-Sagi, K., Wulfmeyer, V., and Yiou, P. (2013), "The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project," *Climate Dynamics*, 41, 2555–2575.

Wadsworth, J. L. and Tawn, J. A. (2012), "Dependence modelling for spatial extremes," *Biometrika*, 1, 1–20.

— (2014), "Efficient inference for spatial extreme value processes associated to log-Gaussian random functions," *Biometrika*, 101, 1–15.

Wand, M. P. (1992), "Error analysis for general multivariate kernel estimators," *J. Nonparamtr. Statist.*, 2, 1–15.

Wand, M. P. and Jones, M. C. (1994), "Multivariate plug-in bandwidth selection," *Comput. Statist.*, 9, 97–116.

— (1995), *Kernel Smoothing*, vol. 60, London: Chapman and Hall.

Wang, Y. and Stoev, S. A. (2010), "On the structure and representations of max-stable processes," *Adv. in Appl. Probab.*, 42, 855–877.

Wood, A. T. and Chan, G. (1994), "Simulation of stationary Gaussian processes in [0, 1] d," *Journal of computational and graphical statistics*, 3, 409–432.

Zhang, H. and El-Shaarawi, A. (2010), "On spatial skew-Gaussian processes and applications," *Environmetrics*, 21, 33–47.

Zhu, X., Bowman, K. P., and Genton, M. G. (2014), "Incorporating geostrophic wind information for improved space-time short-term wind speed forecasting," *Ann. Appl. Stat.*, 8, 1782–1799.